GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: February 27, 2003, 14:12:34 ; Search time 53 Seconds
(without alignments)
6793.180 Million cell updates/sec

Title: US-10-007-521-11
Perfect score: 1174
Sequence: 1 GAGCAGCACCCCTCAAGCTG.....AAAAAAAAAAAAAAAAAAAA 1174

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1174	100.0	1174	2	US-08-872-437-1	Sequence 1, Appli
2	1174	100.0	1174	3	US-08-651-136C-11	Sequence 11, Appl
3	1174	100.0	1174	4	US-09-229-911A-11	Sequence 11, Appl
4	451.6	38.5	922	4	US-09-189-060B-65	Sequence 65, Appl
5	437.4	37.3	1060	1	US-08-090-013-1	Sequence 1, Appli
6	437.4	37.3	1060	1	US-08-081-328-1	Sequence 1, Appli
7	437.4	37.3	1060	1	US-08-232-249-1	Sequence 1, Appli
8	437.4	37.3	1060	2	US-08-921-426-7	Sequence 7, Appli
9	437.4	37.3	1060	2	US-08-833-642A-1	Sequence 1, Appli
10	437.4	37.3	1060	2	US-08-140-008A-3	Sequence 3, Appli
11	437.4	37.3	1060	2	US-08-389-423-1	Sequence 1, Appli
12	437.4	37.3	1060	3	US-08-816-915-7	Sequence 7, Appli
13	437.4	37.3	1060	4	US-09-189-060B-55	Sequence 55, Appl
14	437.4	37.3	1060	4	US-09-230-665-1	Sequence 1, Appli
15	437.4	37.3	1060	4	US-09-189-028-1	Sequence 1, Appli
16	437.4	37.3	1060	5	PCT-US95-07743-7	Sequence 7, Appli
17	411.4	35.0	919	4	US-09-189-060B-67	Sequence 67, Appl
18	392.2	33.4	807	4	US-09-230-665-5	Sequence 5, Appli
19	392	33.4	922	4	US-09-189-060B-73	Sequence 73, Appl
20	377.2	32.1	1257	3	US-09-230-222-2	Sequence 2, Appli
21	375.6	32.0	1257	4	US-09-230-225B-5	Sequence 5, Appli
22	369.8	31.5	1154	3	US-08-651-136C-7	Sequence 7, Appli
23	369.8	31.5	1154	4	US-09-229-911A-7	Sequence 7, Appli
24	364.6	31.1	924	4	US-09-189-060B-69	Sequence 69, Appl
25	364	31.0	927	4	US-09-254-733-4	Sequence 4, Appli
26	353.6	30.1	936	4	US-09-329-350-30	Sequence 30, Appl
27	337	28.7	894	3	US-08-651-136C-3	Sequence 3, Appli
28	337	28.7	894	4	US-09-229-911A-3	Sequence 3, Appli
29	336.2	28.6	912	4	US-09-189-060B-71	Sequence 71, Appl
30	334.8	28.5	927	3	US-08-651-136C-5	Sequence 5, Appli
31	334.8	28.5	927	4	US-09-229-911A-5	Sequence 5, Appli
32	315.2	26.8	1473	1	US-08-090-013-3	Sequence 3, Appli
33	315.2	26.8	1473	1	US-08-081-328-3	Sequence 3, Appli
34	315.2	26.8	1473	1	US-08-232-249-3	Sequence 3, Appli
35	315.2	26.8	1473	2	US-08-833-642A-3	Sequence 3, Appli
36	315.2	26.8	1473	2	US-08-389-423-3	Sequence 3, Appli
37	315.2	26.8	1473	4	US-09-230-665-3	Sequence 3, Appli
38	315.2	26.8	1473	4	US-09-189-028-3	Sequence 3, Appli
39	313.6	26.7	885	3	US-08-651-136C-23	Sequence 23, Appl
40	313.6	26.7	885	4	US-09-229-911A-23	Sequence 23, Appl
41	302	25.7	1132	3	US-08-651-136C-21	Sequence 21, Appl
42	302	25.7	1132	4	US-09-229-911A-21	Sequence 21, Appl
43	288.2	24.5	1423	3	US-08-651-136C-9	Sequence 9, Appli
44	288.2	24.5	1423	4	US-09-229-911A-9	Sequence 9, Appli
45	286.2	24.4	984	4	US-09-230-665-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-872-437-1
; Sequence 1, Application US/08872437
; Patent No. 5958082
; GENERAL INFORMATION:
; APPLICANT: Lund, Henrik
; APPLICANT: Kalum, Lisbeth
; TITLE OF INVENTION: Garments With Considerable Variation In
; TITLE OF INVENTION: Abrasion Level
; FILE REFERENCE: 4888.200-US
; CURRENT APPLICATION NUMBER: US/08/872,437
; CURRENT FILING DATE: 1997-06-10
; EARLIER APPLICATION NUMBER: 1276/96
; EARLIER FILING DATE: 1996-11-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1174
; TYPE: DNA
; ORGANISM: Thielavia terrestris
US-08-872-437-1

Query Match 100.0%; Score 1174; DB 2; Length 1174;
Best Local Similarity 100.0%; Pred. No. 5.7e-224;
Matches 1174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60

Qy      61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120

Qy     121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180

Qy     181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240

Qy     241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300

Qy     301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360
```

```
Qy     361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420

Qy     421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480

Qy     481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540

Qy     541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600

Qy     601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660

Qy     661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720

Qy     721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780

Qy     781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840

Qy     841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900

Qy     901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960

Qy     961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020

Qy    1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080

Qy    1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAA 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAA 1140

Qy    1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
           ||||||||||||||||||||||||||||||||||
Db    1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174


RESULT 2
US-08-651-136C-11
; Sequence 11, Application US/08651136C
; Patent No. 6001639
;  GENERAL INFORMATION:
;    APPLICANT:  Schulein, Martin
;    APPLICANT:  Andersen, Lene N.
;    APPLICANT:  Lassen, Soren F.
;    APPLICANT:  Kauppinen, Markus S.
;    APPLICANT:  Lange, Lene
;    APPLICANT:  Nielsen, Ruby I.
;    APPLICANT:  Ihara, Michiko
;    APPLICANT:  Takagi, Shinobu
;    TITLE OF INVENTION:  No. 6001639el Endoglucanases
;    NUMBER OF SEQUENCES:  109
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6001639o No. 6001639disk of No. 6001639th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,136C
;      FILING DATE:  21-MAY-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4366.200-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1174 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  60..956
US-08-651-136C-11

  Query Match             100.0%;  Score 1174;  DB 3;  Length 1174;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-224;
  Matches 1174;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60

Qy      61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120

Qy     121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180

Qy     181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240

Qy     241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300

Qy     301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360

Qy     361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420

Qy     421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480

Qy     481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540
```

```
Qy    541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600

Qy    601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660

Qy    661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720

Qy    721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780

Qy    781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840

Qy    841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900

Qy    901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960

Qy    961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020

Qy   1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080

Qy   1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAAA 1140

Qy   1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
          ||||||||||||||||||||||||||||||||||
Db   1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174


RESULT 3
US-09-229-911A-11
; Sequence 11, Application US/09229911A
; Patent No. 6387690
;   GENERAL INFORMATION:
;        APPLICANT: Schulein, Martin
;                   Andersen, Lene N.
;                   Lassen, Soren F.
;                   Kauppinen, Markus S.
;                   Lange, Lene
;                   Nielsen, Ruby I.
;                   Ihara, Michiko
;                   Takagi, Shinobu
;        TITLE OF INVENTION: No. 6387690el Endoglucanases
;        NUMBER OF SEQUENCES: 109
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: No. 6387690o No. 6387690disk of No. 6387690th America, Inc.
;             STREET: 405 Lexington Avenue, 64th Floor
;             CITY: New York
;             STATE: New York
;             COUNTRY: United States of America
;             ZIP: 10174-6401
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/229,911A
;             FILING DATE: 13-Jan-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/651,136
;             FILING DATE: 21-MAY-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Lambiris, Elias J.
;             REGISTRATION NUMBER: 33,728
;             REFERENCE/DOCKET NUMBER: 4366.200-US
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 212-867-0123
;             TELEFAX: 212-878-9655
;   INFORMATION FOR SEQ ID NO: 11:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1174 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  60..956
;        SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-229-911A-11

  Query Match             100.0%;  Score 1174;  DB 4;  Length 1174;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-224;
  Matches 1174;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60

Qy     61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120

Qy    121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180

Qy    181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240

Qy    241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300

Qy    301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360

Qy    361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420

Qy    421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480

Qy    481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540

Qy    541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600

Qy    601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660

Qy     661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720

Qy     721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780

Qy     781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840

Qy     841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900

Qy     901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960

Qy     961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020

Qy    1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080

Qy    1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAA 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAA 1140

Qy    1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
           ||||||||||||||||||||||||||||||||||
Db    1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174


RESULT 4
US-09-189-060B-65
; Sequence 65, Application US/09189060B
; Patent No. 6270968
; GENERAL INFORMATION:
;  APPLICANT: Dalboge, Henrik
;  APPLICANT:  Sandal, Thomas
;  APPLICANT:  Kauppinen, Markus
;  APPLICANT:  Borge, Diderichsen
;  TITLE OF INVENTION: Method Of Providing No. 6270968el DNA Sequences
;  FILE REFERENCE: 4772.204-US
;  CURRENT APPLICATION NUMBER: US/09/189,060B
;  CURRENT FILING DATE:  1998-11-10
;  PRIOR APPLICATION NUMBER: PCT/DK97/00216
;  PRIOR FILING DATE: 1997-05-12
;  NUMBER OF SEQ ID NOS: 74
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 65
;   LENGTH: 922
;   TYPE: DNA
;   ORGANISM: Hybrid
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(922)
US-09-189-060B-65

  Query Match           38.5%;  Score 451.6;  DB 4;  Length 922;
  Best Local Similarity  71.6%;  Pred. No. 4.6e-81;
  Matches  630;  Conservative    0;  Mismatches  229;  Indels   21;  Gaps    2;

Qy     102 CTTCCTCTGGTCGCCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGC 161
           ||||   | |  ||  | ||   |  ||    ||||  ||||||| | ||||||| |||
Db      43 CTTCAGGTCGCGGCACCTGCTTTCGCTGCTGATGGCAGGTCCACGCGGTACTGGGATTGC 102

Qy     162 TGCAAGCCGTCGTGCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGC 221
           || |||||||||||| | |||||||| |||||   ||| | ||| || ||||||||  ||
Db     103 TGTAAGCCGTCGTGCTCGTGGCCCGGCAAGGCGCTCGTGAACCAGCCCGTCTACGCCCGC 162

Qy     222 GATGCCAACTTCCAGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCG 281
           | || |||||||||||| |  |||||  ||| | | |||| |||||| | ||||||||
Db     163 AACGCAAACTTCCAGCGCATCACCGACCCCAACGCCAAGTCCGGCTGCGATGGCGGCTCC 222

Qy     282 GCCTACTCCTGCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTC 341
           |||| |||||||||||||||| || ||||| |||| |||| |  | ||||||| |||
Db     223 GCCTTCTCCTGCGCCGACCAGACCCCGTGGGCCGTGAGCGACGACTTTGCCTACGGTTTC 282

Qy     342 GCCGCGACGAGCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTC 401
           || || |||    |||||||  |||||| || |||||||||| ||||||||  |||
Db     283 GCGGCTACGGCGCTCGCCGGCCAGTCCGAGTCTTCGTGGTGCTGTGCCTGCTACGAACTC 342

Qy     402 ACCTTCACTTCCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGC 461
           |||||||||| || ||||| || |||||||  ||||  || ||||| || |||||||||
Db     343 ACCTTCACTTCGGGCCCCGTTGCTGGCAAGAAGATGGCTGTCCAGTCCACCAGCACTGGC 402

Qy     462 GGCGACCTGGGAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTC 521
           || ||||| || || |||| || || ||  ||  |||||| || ||||| || |||||||||
Db     403 GGTGACCTCGGTAGCAACCACTTTGACCTCAACATGCCAGGTGGCGGTGTCGGCATCTTC 462

Qy     522 AACGGCTGCAGCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCG 581
           |||||||||    | ||| | ||||| ||| |||||    |  || |||||| | |||||
Db     463 GACGGCTGCTCGCCTCAGGTTGGCGGTCTCGCCGGCCAGCGCTATGGCGGCGTCTCGTCC 522

Qy     582 CGCGACCAGTGCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGAC 641
           |||  | | ||||| |||||||||||| | ||||||||||||||| | ||||| |  |||
Db     523 CGCAGCGAATGCGACTCCTTCCCCGCGGCACTCAAGCCCGGCTGCTACTGGCGCTACGAC 582

Qy     642 TGGTTCCAGAACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAG 701
           |||||  |||||||||||||| ||||  ||||  ||||  ||||| |||||||| ||||||
Db     583 TGGTTTAAGAACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAG 642

Qy     702 ATCGTTGCCCGCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCC 761
           ||||  || ||| |||| |||   ||||||||||||  || |||||| | | ||     |
Db     643 CTCGTCGCTCGCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATC 702

Qy     762 CCAAGCGGTGGCAACGG----------------TGGCACCGGGACGCCCACGTCGACT 803
           ||   | |  ||| | |                |  || |   |   ||||||| ||
Db     703 CCCTCCAGCAGCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACC 762

Qy     804 GCGCCTGGGTCGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAG 860
           |  |    ||| ||| | |      ||         ||| |||||||  || ||| |
Db     763 TCCACCACCTCGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGG 822

Qy     861 TGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACC 920
           |||||||||||||| ||||  ||||  ||||| |||||||||| ||| |||||| |||
Db     823 TGGGCTCAGTGCGGCGGCAATGGCTGGAGCGGGTGCACCACCTGCGTCGCTGGCAGCACT 882

Qy     921 TGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960
           |||  |||| | || ||||  |||   |||||||| || |
Db     883 TGCACGAAGATTAATGACTGGTACCATCAGTGCCTGTAGA 922


RESULT 5
US-08-090-013-1
; Sequence 1, Application US/08090013
; Patent No. 5443750
;  GENERAL INFORMATION:
;    APPLICANT:  CONVENTS, ANDRE C
;    APPLICANT:  BUSCH, ALFRED
;    APPLICANT:  BAECK, ANDRE C
;    TITLE OF INVENTION:  DETERGENT COMPOSITION WITH HIGH ACTIVITY
;    TITLE OF INVENTION:  CELLULASE AND SOFTENING CLAYS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THE PROCTER & GAMBLE COMPANY
;      STREET:  5299 SPRING GROVE AVENUE
;      CITY:  CINCINNATI
;      STATE:  OHIO
```

```
;      COUNTRY:  USA
;      ZIP:  45217
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/090,013
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 91202880.0
;      FILING DATE:  06-NOV-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PATEL, KEN K.
;      REGISTRATION NUMBER:  33,988
;      REFERENCE/DOCKET NUMBER:  CM393
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  513-627-6437
;      TELEFAX:  513-627-4854
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10..924
US-08-090-013-1

  Query Match             37.3%;  Score 437.4;  DB 1;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3e-78;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | |||||| | | | |||| ||||||  |    |||||    ||||||||  |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||||| || ||||| |||||||||  |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||| |    ||  |    ||||||||||||||||||| ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           |||||||  || || |||||||  ||||| || ||||| || ||||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| |||||||||||||| |||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
             | |||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | |||||||||||||| | ||||| || |||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| |||| |||| |||| ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   |||||||||||  || |||||| | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                 |  || |   |   ||||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||         ||| ||||||||  || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| ||||  ||||  ||||| |||||||||| ||| |||||| ||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || ||||  |||   |||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 6
US-08-081-328-1
; Sequence 1, Application US/08081328
; Patent No. 5520838
;  GENERAL INFORMATION:
;    APPLICANT:  BAECK, ANDRE C.
;    APPLICANT:  CEULEMANS, RAPHAEL ANGELINE A.
;    APPLICANT:  BUSCH, ALFRED (NMN)
;    TITLE OF INVENTION:  COMPACT DETERGENT COMPOSITIONS WITH
;    TITLE OF INVENTION:  HIGH ACTIVITY CELLULASE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THE PROCTER & GAMBLE COMPANY
;      STREET:  11810 East Miami River Road
;      CITY:  CINCINNATI
;      STATE:  OHIO
;      COUNTRY:  USA
;      ZIP:  45253-8707
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/081,328
;      FILING DATE:  11/19/93
;      CLASSIFICATION:  252
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ZERBY, KIM WILLIAM
;      REGISTRATION NUMBER:  32,323
;      REFERENCE/DOCKET NUMBER:  CM356M
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  513-627-2885
;      TELEFAX:  513-627-0318
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10..924
US-08-081-328-1
```

```
  Query Match             37.3%;  Score 437.4;  DB 1;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3e-78;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |    |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      |||||| | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||    | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | ||||||  | | | |||| ||||||  |    |||||    ||||||||  |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||||  || ||||| |||||||||  |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
           || |||||| |    ||  |    ||||||||||||||||||| ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           |||||||  || || |||||||  ||||| || ||||| || |||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| |||||||||||||| |||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
              |||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | ||||||||||||||| | ||||| || |||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| |||| |||| |||| ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   |||||||||||  || |||||| | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                   |  || |   |   ||||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||          ||| ||||||||  || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| ||||  ||||  ||||| |||||||||| ||| ||||||  ||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || ||||  |||   |||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 7
US-08-232-249-1
; Sequence 1, Application US/08232249
; Patent No. 5610129
;  GENERAL INFORMATION:
;    APPLICANT:  McCORQUODALE, FINLAY (NMN)
;    APPLICANT:  BUSCH, ALFRED (NMN)
;    TITLE OF INVENTION:  DYE TRANSFER INHIBITING COMPOSITIONS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THE PROCTER & GAMBLE COMPANY
;      STREET:  5299 SPRING GROVE AVENUE
;      CITY:  CINCINNATI
;      STATE:  OHIO
;      COUNTRY:  USA
;      ZIP:  45217
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,249
;      FILING DATE:  05-MAY-1994
;      CLASSIFICATION:  510
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 91202882.6
;      FILING DATE:  06-NOV-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ALLEN, GEORGE W.
;      REGISTRATION NUMBER:  26,143
;      REFERENCE/DOCKET NUMBER:  CM-395
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  513-627-5946
;      TELEFAX:  513-627-8118
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10..924
US-08-232-249-1

  Query Match             37.3%;  Score 437.4;  DB 1;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3e-78;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |    |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      |||||| | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||    | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | ||||||  | | | |||| ||||||  |    |||||    ||||||||  |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||||  || ||||| |||||||||  |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
           || |||||| |    ||  |    ||||||||||||||||||| ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           |||||||  || || |||||||  ||||| || ||||| || |||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418
```

```
Qy    472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGGTCGTGGGCATCTTCAACGGCTGCA 531
          | || ||||| |||||| ||  ||| ||||||||||||||| |||||||| |||| ||||
Db    419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy    532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
            | ||||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db    479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy    592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
          |||||   |||||||   | ||||||||||||||| | ||||| || |||||||||  |||
Db    539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy    652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
          |||||||||| ||||  ||||  ||||  ||||| |||||||| |||||| |||| || |
Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
          || |||| |||   |||||||||||  || |||||| | | ||     |||   | |
Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy    772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
          ||| | |                 |  || |   |   ||||||| ||  |  |    |
Db    719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy    814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
          || ||| | |       ||         ||| ||||||||  || ||| ||||||||||
Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
          |||| ||||  ||||  ||||| |||||||||| ||| |||||| ||| |||  ||||
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
          | || ||||  |||   |||||||| || ||
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 8
US-08-921-426-7
; Sequence 7, Application US/08921426
; Patent No. 5837847
;  GENERAL INFORMATION:
;    APPLICANT:  Royer, John C
;    APPLICANT:  Moyer, Donna L
;    APPLICANT:  Yoder, Wendy T
;    APPLICANT:  Shuster, Jeffrey R
;    TITLE OF INVENTION:  NON-TOXIC, NON-TOXIGENIC, NON-PATHOGENIC
;    TITLE OF INVENTION:  FUSARIUM EXPRESSION SYSTEM
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5837847o No. 5837847disk of No. 5837847th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/921,426
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/456,433
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  US 08/404,678
;      FILING DATE:  15-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris Dr., Cheryl H.
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  4216.010-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10..924
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  73..924
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  10..72
US-08-921-426-7

  Query Match             37.3%;  Score 437.4;  DB 2;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3e-78;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy     55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
          ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db      5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy    115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
          ||   || ||   ||||||      ||||||  | ||||||||||||||||||| ||||
Db     65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy    175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
          |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db    119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy    235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
          ||||  |  | |||||| | | | |||| |||||| |    |||||    |||||||| |
Db    179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy    292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
          |||||||||||| || ||||| |||||||| |  ||||   ||| || || || || ||
Db    239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy    352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
          || |||||| |    ||  |    ||||||||||||||||||| |||||||||||  |
Db    299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy    412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
          ||||||| || || |||||||  ||||| || ||||| || ||||||||||| || || |
Db    359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy    472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
          | || ||||| |||||| ||  ||| |||||||||||||| |||||||| |||| ||||
Db    419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy    532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
            | ||||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db    479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy    592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
          |||||   |||||||   | ||||||||||||||| | ||||| || |||||||||  |||
Db    539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy    652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
          |||||||||| ||||  ||||  ||||  ||||| |||||||| |||||| |||| || |
Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
          || |||| |||   |||||||||||  || |||||| | | ||     |||   | |
Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718
```

```
Qy     772 GCAACGG------------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                  |  || |   |   ||||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||        ||| ||||||||  || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| ||||  ||||  ||||| |||||||||||| ||| |||||| ||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || ||||  |||   |||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 9
US-08-833-642A-1
; Sequence 1, Application US/08833642A
; Patent No. 5883066
;  GENERAL INFORMATION:
;    APPLICANT:  Ivan M. A. J. Herbots et al.
;    TITLE OF INVENTION:  Liquid Detergent Compositions
;    TITLE OF INVENTION:  Containing Cellulase and Amine
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jackie Ann Zurcher
;      ADDRESSEE:  Dinsmore & Shohl LLP
;      STREET:  255 E. Fifth Street
;      STREET:  1900 Chemed Center
;      CITY:  Cincinnati
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  45202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/833,642A
;      FILING DATE:  April 8, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Zurcher, J. A.
;      REGISTRATION NUMBER:  P42,251
;      REFERENCE/DOCKET NUMBER:  CM551C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (513) 977-8377
;      TELEFAX:  (513) 977-8141
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Humicola insolens
;      STRAIN:  DSM 1800
;    FEATURE:
;      NAME/KEY:  mat peptide
;      LOCATION:  73.927
;    FEATURE:
;      NAME/KEY:  sig peptide
;      LOCATION:  10.72
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10.927
US-08-833-642A-1


  Query Match             37.3%;  Score 437.4;  DB 2;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3e-78;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |    |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||||  ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||    | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | ||||||  | | | |||| ||||||  |    |||||    ||||||||  |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           ||||||||||||  || ||||| |||||||||  |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||  |    ||  |  ||||||||||||||||||||  ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           |||||||  || ||  |||||||  |||||  || |||||  || ||||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || |||||  ||||||  ||  ||| ||||||||||||||  |||||||| ||||  ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
             | ||||||||||| || ||||||   |  |||||||||| |||||  ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | ||||||||||||||| | |||||  || |||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| ||||  ||||  ||||  ||||| ||||||||  ||||||  |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   ||||||||||||  || |||||| | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG------------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                  |  || |   |   ||||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||        ||| ||||||||  || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| ||||  ||||  ||||| |||||||||||| ||| |||||| ||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || ||||  |||   |||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 10
US-08-140-008A-3
; Sequence 3, Application US/08140008A
; Patent No. 5914306
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  SVENDSEN, Allan
;    APPLICANT:  VON DER OSTEN, Claus
;    APPLICANT:  CLAUSEN, Ib Groth
;    APPLICANT:  PATKAR, Shamkant Anant
;    APPLICANT:  BORCH, Kim
;    TITLE OF INVENTION:  STABILIZED ENZYMES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5914306o No. 5914306disk of No. 5914306th America
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/140,008A
;      FILING DATE:  22-OCT-1993
;      CLASSIFICATION:  252
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  3601.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 867-0123
;      TELEFAX:  (212) 878-9655
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10..924
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  10..72
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  73..924
US-08-140-008A-3

  Query Match             37.3%;  Score 437.4;  DB 2;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3e-78;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | |||||| | | | |||| ||||||  |     |||||    ||||||||  |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||| || ||||| |||||||| |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
           || |||||| |    ||  |    |||||||||||||||||| ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           ||||||| || || |||||||  ||||| || ||||| || |||||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| |||||||||||||| |||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
           |  |||||||||| || ||||||    |  ||||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | ||||||||||||||| | ||||| || |||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| |||| |||| |||| ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   ||||||||||||  || |||||| | | ||      |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                  |  || |   |   ||||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||        ||| ||||||||  || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| |||| |||| ||||| |||||||||| ||| |||||| ||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || ||||  |||   |||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 11
US-08-389-423-1
; Sequence 1, Application US/08389423
; Patent No. 5948672
;  GENERAL INFORMATION:
;    APPLICANT:  Rasmussen, Grethe
;    APPLICANT:  Mikkelsen, Jan Moller
;    APPLICANT:  Schulein, Martin
;    APPLICANT:  Patkar, Shankant A.
;    APPLICANT:  Hagen, Fred
;    TITLE OF INVENTION:  A Cellulase Preparation Comprising an
;    TITLE OF INVENTION:  Endoglucanase Enzyme
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5948672o No. 5948672disk of No. 5948672th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/389,423
;      FILING DATE:  14-FEB-1995
```

```
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lambiris, Elias J.
;       REGISTRATION NUMBER:  33,728
;       REFERENCE/DOCKET NUMBER:  3469.214-US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-867-0123
;       TELEFAX:  212-878-9655
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1060 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Humicola insolens
;       STRAIN:  DSM 1800
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  73..924
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  10..72
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  10..924
US-08-389-423-1

  Query Match             37.3%;  Score 437.4;  DB 2;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3e-78;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||    | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | ||||||  | | | ||||  |||||| |    |||||    |||||||| |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||| || ||||| |||||||| |  ||||   ||| || || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||| |    ||  |    |||||||||||||||||| |||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           |||||||  || ||  |||||||  |||||  || |||||  || |||||||||||  || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| ||||| ||  ||| |||||||||||| |||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
             | ||||||||||  || ||||||    |  ||||||||||  |||||  |||  ||  |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | ||||||||||||||| | |||||  || |||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| ||||  ||||  ||||  ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   |||||||||||  || ||||||  | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG------------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                  |  || |   |   ||||||| ||  |  |   |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||         ||| ||||||||  || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| ||||  ||||  ||||| |||||||||| ||| |||||| ||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || ||||  |||   ||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 12
US-08-816-915-7
; Sequence 7, Application US/08816915
; Patent No. 6060305
;  GENERAL INFORMATION:
;    APPLICANT:  Royer, John C
;    APPLICANT:  Moyer, Donna L
;    APPLICANT:  Yoder, Wendy T
;    APPLICANT:  Shuster, Jeffrey R
;    TITLE OF INVENTION:  NON-TOXIC, NON-TOXIGENIC, NON-PATHOGENIC
;    TITLE OF INVENTION:  FUSARIUM EXPRESSION SYSTEM
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6060305o No. 6060305disk of No. 6060305th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/816,915
;      FILING DATE:  13-MAR-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris Dr., Cheryl H.
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  4216.240-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10..924
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  73..924
```

```
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  10..72
US-08-816-915-7

  Query Match             37.3%;  Score 437.4;  DB 3;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3e-78;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | |||||| | | | |||| |||||| |    |||||    |||||||| |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||||  || ||||| |||||||| |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||| |    ||  |    ||||||||||||||||||| ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           ||||||| || || ||||||| ||||| || ||||| || |||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| ||||||||||||| |||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
           | ||||||||||| || ||||||    |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | ||||||||||||||| | ||||| || |||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| |||  |||  ||||  ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   ||||||||||||  || |||||| | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                 |  || |   |  ||||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||          ||| ||||||||  || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| ||||  ||||  ||||| |||||||||||| ||| |||||| ||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || ||||  |||   ||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 13
US-09-189-060B-55
; Sequence 55, Application US/09189060B
; Patent No. 6270968
; GENERAL INFORMATION:
;  APPLICANT: Dalboge, Henrik
;  APPLICANT:  Sandal, Thomas
;  APPLICANT:  Kauppinen, Markus
;  APPLICANT:  Borge, Diderichsen
;  TITLE OF INVENTION: Method Of Providing No. 6270968el DNA Sequences
;  FILE REFERENCE: 4772.204-US
;  CURRENT APPLICATION NUMBER: US/09/189,060B
;  CURRENT FILING DATE:  1998-11-10
;  PRIOR APPLICATION NUMBER: PCT/DK97/00216
;  PRIOR FILING DATE: 1997-05-12
;  NUMBER OF SEQ ID NOS: 74
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 55
;   LENGTH: 1060
;   TYPE: DNA
;   ORGANISM: Humicola insolens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10)...(927)
;   NAME/KEY: mat_peptide
;   LOCATION: (73)...(927)
;   NAME/KEY: sig_peptide
;   LOCATION: (10)...(72)
US-09-189-060B-55

  Query Match             37.3%;  Score 437.4;  DB 4;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3e-78;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | |||||| | | | |||| |||||| |    |||||    |||||||| |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||||  || ||||| |||||||| |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||| |    ||  |    ||||||||||||||||||| ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           ||||||| || || ||||||| ||||| || ||||| || |||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| ||||||||||||| |||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
           | ||||||||||| || ||||||    |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | ||||||||||||||| | ||||| || |||||||| |||
```

```
Db    539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy    652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
          |||||||||| ||||  ||||  ||||  ||||| |||||||| ||||||  |||| || |
Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
          || |||| |||   ||||||||||||  || ||||||  | | ||     |||   | |
Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy    772 GCAACGG------------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
          |||  | |                  |  || |   |   ||||||| ||  |  |    |
Db    719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy    814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
          || ||| | |       ||         ||| ||||||||  || ||| ||||||||||
Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
          |||| ||||  ||||  ||||| |||||||||| ||| |||||| ||| |||  ||||
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
          | || ||||  |||   |||||||| || ||
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 14
US-09-230-665-1
; Sequence 1, Application US/09230665
; Patent No. 6322595
; GENERAL INFORMATION:
;  APPLICANT: Boyer, Stanton L
;  TITLE OF INVENTION: Detergent Composition Comprising Two Cellulase
;  TITLE OF INVENTION: Components, with and without a Cellulose-Binding Domain
;  FILE REFERENCE: 6191 Sequence Listing (8 Sequences)
; Patent No. 6322595
;  CURRENT APPLICATION NUMBER: US/09/230,665
;  CURRENT FILING DATE: 1999-01-29
;  EARLIER APPLICATION NUMBER: 60/023,125
;  EARLIER FILING DATE: 1996-07-30
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 1060
;   TYPE: DNA
;   ORGANISM: Humicola insolens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10)..(924)
;   FEATURE:
;   NAME/KEY: mat_peptide
;   LOCATION: (73)..(924)
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: (10)..(72)
US-09-230-665-1

  Query Match             37.3%;  Score 437.4;  DB 4;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3e-78;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy     55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
          ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db      5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy    115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
          ||   || ||   ||||||      ||||||  | |||||||||||||||||| ||||
Db     65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy    175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
          |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db    119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy    235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
          ||||  |  | ||||||  | | | |||| ||||||  |    |||||    ||||||||  |
Db    179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy    292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
          |||||||||||| || |||| |||||||| |  ||||   ||| || || || || ||
Db    239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy    352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
            || |||||| |    ||  |  ||||||||||||||||||| ||||||||||| |
Db    299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy    412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
          ||||||| || || |||||||  ||||| || ||||| || |||||||||| || || |
Db    359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy    472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
          | || ||||| |||||| ||  ||| |||||||||||||| |||||||| |||| ||||
Db    419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy    532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
          | ||||||||||| || ||||||   |  |||||||||| |||| ||| || |||
Db    479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy    592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
          |||||   |||||||   | |||||||||||||| | |||| || |||||||| |||
Db    539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy    652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
          |||||||||| |||| |||| |||| |||||| |||||||| |||||| |||| || |
Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
          || |||| |||   ||||||||||||  || ||||||  | | ||     |||   | |
Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy    772 GCAACGG------------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
          |||  | |                  |  || |   |   ||||||| ||  |  |    |
Db    719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy    814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
          || ||| | |       ||         ||| ||||||||  || ||| ||||||||||
Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
          |||| ||||  ||||  ||||| |||||||||| ||| |||||| ||| |||  ||||
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
          | || ||||  |||   |||||||| || ||
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 15
US-09-189-028-1
; Sequence 1, Application US/09189028
; Patent No. 6423524
;  GENERAL INFORMATION:
;    APPLICANT:  Rasmussen, Grethe
;    APPLICANT:  Mikkelsen, Jan Moller
;    APPLICANT:  Schulein, Martin
;    APPLICANT:  Patkar, Shankant A.
;    APPLICANT:  Hagen, Fred
;    TITLE OF INVENTION:  A Cellulase Preparation Comprising an
;    TITLE OF INVENTION:  Endoglucanase Enzyme
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6423524o No. 6423524disk of No. 6423524th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
```

```
;       ZIP:  10174-6401
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/189,028
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/389,423
;       FILING DATE:  14-FEB-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lambiris, Elias J.
;       REGISTRATION NUMBER:  33,728
;       REFERENCE/DOCKET NUMBER:  3469.214-US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-867-0123
;       TELEFAX:  212-878-9655
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1060 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Humicola insolens
;       STRAIN:  DSM 1800
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  73..924
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  10..72
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  10..924
US-09-189-028-1

  Query Match             37.3%;  Score 437.4;  DB 4;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3e-78;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||    | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | |||||| | | | |||| |||||| |    |||||    |||||||| |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||||| || ||||| |||||||||  |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||| |    ||  |    ||||||||||||||||||| ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           ||||||| || || |||||||  ||||| || ||||| || |||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| |||||||||||||| |||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
             | |||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | |||||||||||||| | ||||| || ||||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| |||| |||| |||| ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   |||||||||||  || |||||| | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                 |  || |   |   ||||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||         ||| ||||||||  || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| |||| |||| ||||| |||||||||| ||| |||||| ||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || |||| |||   |||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


Search completed: February 27, 2003, 17:47:34
Job time : 59 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: February 27, 2003, 17:47:40 ; Search time 3182 Seconds
(without alignments)
10737.490 Million cell updates/sec

Title: US-10-007-521-11
Perfect score: 1174
Sequence: 1 GAGCAGCACCCCTCAAGCTG.....AAAAAAAAAAAAAAAAAAAA 1174

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1174	100.0	1174	6	A92311	A92311 Sequence 1
2	1174	100.0	1174	6	AR075389	AR075389 Sequence
3	1174	100.0	1174	6	AR094310	AR094310 Sequence
4	451.6	38.5	922	6	A68070	A68070 Sequence 65
5	451.6	38.5	922	6	AR163167	AR163167 Sequence
6	437.4	37.3	1060	6	A21793	A21793 H.insolens
7	437.4	37.3	1060	6	A23635	A23635 H. insolens
8	437.4	37.3	1060	6	A23644	A23644 H. insolens
9	437.4	37.3	1060	6	A23953	A23953 Endoglucana
10	437.4	37.3	1060	6	A23957	A23957 Endoglucana
11	437.4	37.3	1060	6	A41658	A41658 Sequence 1
12	437.4	37.3	1060	6	A68060	A68060 Sequence 55
13	437.4	37.3	1060	6	AR059002	AR059002 Sequence
14	437.4	37.3	1060	6	AR072921	AR072921 Sequence
15	437.4	37.3	1060	6	AR163160	AR163160 Sequence
16	437.4	37.3	1060	6	BD010852	BD010852 Cellulase
17	437.4	37.3	1060	6	I13884	I13884 Sequence 1
18	437.4	37.3	1060	6	I21316	I21316 Sequence 1
19	437.4	37.3	1060	6	I57983	I57983 Sequence 1
20	428.6	36.5	1057	6	BD002248	BD002248 Cellulase
21	411.4	35.0	919	6	A68072	A68072 Sequence 67
22	411.4	35.0	919	6	AR163168	AR163168 Sequence
23	392	33.4	922	6	A68078	A68078 Sequence 73
24	392	33.4	922	6	AR163171	AR163171 Sequence
25	377.2	32.1	1257	6	AR121354	AR121354 Sequence
26	374.6	31.9	1904	8	AB003107	AB003107 Humicola
27	369.8	31.5	1154	6	AR094308	AR094308 Sequence
28	364.6	31.1	924	6	A68074	A68074 Sequence 69
29	364.6	31.1	924	6	AR163169	AR163169 Sequence
30	337	28.7	894	6	AR094306	AR094306 Sequence
31	336.2	28.6	912	6	A68076	A68076 Sequence 71
32	336.2	28.6	912	6	AR163170	AR163170 Sequence
33	334.8	28.5	927	6	AR094307	AR094307 Sequence
34	315.2	26.8	1443	8	FSOKCH	L29381 Fusarium ox
35	315.2	26.8	1473	6	A21795	A21795 F.oxysporum
36	315.2	26.8	1473	6	A23637	A23637 F. oxysporu
37	315.2	26.8	1473	6	A23646	A23646 F. oxysporu
38	315.2	26.8	1473	6	A23955	A23955 Endoglucana
39	315.2	26.8	1473	6	A23959	A23959 Endoglucana
40	315.2	26.8	1473	6	A41660	A41660 Sequence 3
41	315.2	26.8	1473	6	AR072922	AR072922 Sequence
42	315.2	26.8	1473	6	BD002249	BD002249 Cellulase
43	315.2	26.8	1473	6	BD010853	BD010853 Cellulase
44	315.2	26.8	1473	6	I13885	I13885 Sequence 3
45	315.2	26.8	1473	6	I21317	I21317 Sequence 3

ALIGNMENTS

RESULT 1
A92311
LOCUS A92311 1174 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent EP0843041.
ACCESSION A92311
VERSION A92311.1 GI:6741083
KEYWORDS .
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 1174)
AUTHORS Kalum,L. and Lund,H.
TITLE Garments with considerable variation in abrasion level and process
for its production using cellulolytic enzymes

```
  JOURNAL   Patent: EP 0843041-A 1 20-MAY-1998;
            NOVONORDISK AS (DK)
FEATURES             Location/Qualifiers
     source          1. .1174
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
BASE COUNT      243 a    395 c    320 g    216 t
ORIGIN

  Query Match           100.0%;  Score 1174;  DB 6;  Length 1174;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-197;
  Matches 1174;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60

Qy      61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120

Qy     121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180

Qy     181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240

Qy     241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300

Qy     301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360

Qy     361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420

Qy     421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480

Qy     481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540

Qy     541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600

Qy     601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660

Qy     661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720

Qy     721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780

Qy     781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840

Qy     841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900

Qy     901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960

Qy     961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020

Qy    1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080

Qy    1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAA 1140

Qy    1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
          ||||||||||||||||||||||||||||||||||
Db    1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174


RESULT 2
AR075389
LOCUS       AR075389               1174 bp    DNA     linear   PAT 30-AUG-2000
DEFINITION  Sequence 1 from patent US 5958082.
ACCESSION   AR075389
VERSION     AR075389.1  GI:10002139
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1174)
  AUTHORS   Lund,H. and Kalum,L.
  TITLE     Garments with considerable variation in abrasion level
  JOURNAL   Patent: US 5958082-A 1 28-SEP-1999;
FEATURES             Location/Qualifiers
     source          1. .1174
                     /organism="unknown"
BASE COUNT      243 a    395 c    320 g    216 t
ORIGIN

  Query Match           100.0%;  Score 1174;  DB 6;  Length 1174;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-197;
  Matches 1174;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60

Qy      61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120

Qy     121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180

Qy     181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240

Qy     241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300

Qy     301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360

Qy     361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420
```

```
Qy    421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480

Qy    481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540

Qy    541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600

Qy    601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660

Qy    661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720

Qy    721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780

Qy    781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840

Qy    841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900

Qy    901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960

Qy    961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020

Qy   1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080

Qy   1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAA 1140

Qy   1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
          ||||||||||||||||||||||||||||||||||
Db   1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
```

```
RESULT 3
AR094310
LOCUS       AR094310               1174 bp    DNA     linear   PAT 08-SEP-2000
DEFINITION  Sequence 11 from patent US 6001639.
ACCESSION   AR094310
VERSION     AR094310.1  GI:10021119
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1174)
  AUTHORS   Schulein,M., Andersen,L.Nonboe., Lassen,Sslashedren.Flensted.,
            Kauppinen,M.Sakari., Lange,L., Nielsen,R.Iium., Ihara,M. and
            Takagi,S.
  TITLE     Endoglucanases
  JOURNAL   Patent: US 6001639-A 11 14-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .1174
                     /organism="unknown"
BASE COUNT      243 a    395 c    320 g    216 t
```

ORIGIN

```
  Query Match           100.0%;  Score 1174;  DB 6;  Length 1174;
  Best Local Similarity 100.0%;  Pred. No. 7.5e-197;
  Matches 1174;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60

Qy     61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120

Qy    121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180

Qy    181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240

Qy    241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300

Qy    301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360

Qy    361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420

Qy    421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480

Qy    481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540

Qy    541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600

Qy    601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660

Qy    661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720

Qy    721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780

Qy    781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840

Qy    841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900

Qy    901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960

Qy    961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020
```

```
Qy   1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080

Qy   1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAA 1140

Qy   1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
          ||||||||||||||||||||||||||||||||||
Db   1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
```

RESULT 4
A68070

LOCUS	A68070 922 bp DNA linear PAT 05-MAY-1999
DEFINITION	Sequence 65 from Patent WO9743409.
ACCESSION	A68070
VERSION	A68070.1 GI:4756870
KEYWORDS	.
SOURCE	unidentified.
ORGANISM	unidentified
	unclassified.
REFERENCE	1 (bases 1 to 922)
AUTHORS	Dalboege,H., Diderichsen,B., Sandal,T. and Kauppinen,S.
TITLE	METHOD OF PROVIDING NOVEL DNA SEQUENCES
JOURNAL	Patent: WO 9743409-A 65 20-NOV-1997;
	NOVONORDISK AS (DK)

FEATURES Location/Qualifiers

```
     source          1. .922
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     CDS             <1. .922
                     /note="unnamed protein product"
                     /codon_start=2
                     /protein_id="CAB42308.1"
                     /db_xref="GI:4756871"
                     /translation="PFMMVAWWSLFLYGLQVAAPAFAADGRSTRYWDCCKPSCSWPGK
                     ALVNQPVYARNANFQRITDPNAKSGCDGGSAFSCADQTPWAVSDDFAYGFAATALAGQ
                     SESSWCCACYELTFTSGPVAGKKMAVQSTSTGGDLGSNHFDLNMPGGGVGIFDGCSPQ
                     VGGLAGQRYGGVSSRSECDSFPAALKPGCYWRYDWFKNADNPSFSFRQVQCPAELVAR
                     TGCRRNDDGNFPAVQIPSSSTSSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWA
                     QCGGNGWSGCTTCVAGSTCTKINDWYHQCL"
BASE COUNT      151 a    333 c    262 g    176 t
ORIGIN
```

```
  Query Match             38.5%;  Score 451.6;  DB 6;  Length 922;
  Best Local Similarity   71.6%;  Pred. No. 1e-69;
  Matches  630;  Conservative    0;  Mismatches  229;  Indels   21;  Gaps    2;

Qy    102 CTTCCTCTGGTCGCCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGC 161
          ||||   | |  ||  | ||   |  ||    ||||  ||||||| | ||||||| |||
Db     43 CTTCAGGTCGCGGCACCTGCTTTCGCTGCTGATGGCAGGTCCACGCGGTACTGGGATTGC 102

Qy    162 TGCAAGCCGTCGTGCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGC 221
          || |||||||||||| | |||||||| |||||   ||| | ||| || ||||||||  ||
Db    103 TGTAAGCCGTCGTGCTCGTGGCCCGGCAAGGCGCTCGTGAACCAGCCCGTCTACGCCCGC 162

Qy    222 GATGCCAACTTCCAGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCG 281
           | || |||||||||||| |  |||||  ||| | | |||| |||||| | ||||||||
Db    163 AACGCAAACTTCCAGCGCATCACCGACCCCAACGCCAAGTCCGGCTGCGATGGCGGCTCC 222

Qy    282 GCCTACTCCTGCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTC 341
          |||| |||||||||||||||| || ||||| |||| |||| |   | ||||||| |||
Db    223 GCCTTCTCCTGCGCCGACCAGACCCCGTGGGCCGTGAGCGACGACTTTGCCTACGGTTTC 282

Qy    342 GCCGCGACGAGCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTC 401
          || || |||    |||||||| |||||| || |||||||||| ||||||||||  |||
Db    283 GCGGCTACGGCGCTCGCCGGCCAGTCCGAGTCTTCGTGGTGCTGTGCCTGCTACGAACTC 342

Qy    402 ACCTTCACTTCCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGC 461
          |||||||||||  || ||||| || |||||||  ||||  || ||||| || |||||||||
Db    343 ACCTTCACTTCGGGCCCCGTTGCTGGCAAGAAGATGGCTGTCCAGTCCACCAGCACTGGC 402

Qy    462 GGCGACCTGGGAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTC 521
          || ||||| || || ||||| || ||  ||  |||||| || ||||| || |||||||||
Db    403 GGTGACCTCGGTAGCAACCACTTTGACCTCAACATGCCAGGTGGCGGTGTCGGCATCTTC 462

Qy    522 AACGGCTGCAGCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCG 581
          ||||||||     | ||| | ||||| ||| |||||    |  || |||||| | |||||
Db    463 GACGGCTGCTCGCCTCAGGTTGGCGGTCTCGCCGGCCAGCGCTATGGCGGCGTCTCGTCC 522

Qy    582 CGCGACCAGTGCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGAC 641
          |||  | | ||||| |||||||||||| | ||||||||||||||| | |||||  |  |||
Db    523 CGCAGCGAATGCGACTCCTTCCCCGCGGCACTCAAGCCCGGCTGCTACTGGCGCTACGAC 582

Qy    642 TGGTTCCAGAACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAG 701
          |||||  |||||||||||||| ||||  ||||  ||||  ||||| |||||||| ||||||
Db    583 TGGTTTAAGAACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAG 642

Qy    702 ATCGTTGCCCGCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCC 761
          ||||  || ||| |||| |||   ||||||||||||  || |||||| | | ||     |
Db    643 CTCGTCGCTCGCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATC 702

Qy    762 CCAAGCGGTGGCAACGG-----------------TGGCACCGGGACGCCCACGTCGACT 803
          ||   | |  ||| | |                 |  || |   |  ||||||| ||
Db    703 CCCTCCAGCAGCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACC 762

Qy    804 GCGCCTGGGTCGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAG 860
          | |   ||| ||| | |       ||          ||| ||||||||  || ||| |
Db    763 TCCACCACCTCGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGG 822

Qy    861 TGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACC 920
          |||||||||||||| ||||  ||||  ||||| |||||||||||| ||| |||||| |||
Db    823 TGGGCTCAGTGCGGCGGCAATGGCTGGAGCGGGTGCACCACCTGCGTCGCTGGCAGCACT 882

Qy    921 TGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960
          |||  |||| | || ||||  |||   |||||||| || |
Db    883 TGCACGAAGATTAATGACTGGTACCATCAGTGCCTGTAGA 922
```

RESULT 5
AR163167

LOCUS	AR163167 922 bp DNA linear PAT 17-OCT-2001
DEFINITION	Sequence 65 from patent US 6270968.
ACCESSION	AR163167
VERSION	AR163167.1 GI:16233670
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown.
	Unclassified.
REFERENCE	1 (bases 1 to 922)
AUTHORS	Dalb.o slashed.ge,H., Sandal,T., Kauppinen,M.Sakari. and Diderichsen,Bslashedrge.
TITLE	Method of providing a hybrid polypeptide exhibiting an activity of interest
JOURNAL	Patent: US 6270968-A 65 07-AUG-2001;

FEATURES Location/Qualifiers

```
     source          1. .922
                     /organism="unknown"
BASE COUNT      151 a    333 c    262 g    176 t
ORIGIN
```

```
  Query Match             38.5%;  Score 451.6;  DB 6;  Length 922;
  Best Local Similarity   71.6%;  Pred. No. 1e-69;
  Matches  630;  Conservative    0;  Mismatches  229;  Indels   21;  Gaps    2;

Qy    102 CTTCCTCTGGTCGCCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGC 161
          ||||   | |  ||  | ||   |  ||    ||||  ||||||| | ||||||| |||
Db     43 CTTCAGGTCGCGGCACCTGCTTTCGCTGCTGATGGCAGGTCCACGCGGTACTGGGATTGC 102

Qy    162 TGCAAGCCGTCGTGCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGC 221
          || |||||||||||| | |||||||| |||||   ||| | ||| || ||||||||  ||
Db    103 TGTAAGCCGTCGTGCTCGTGGCCCGGCAAGGCGCTCGTGAACCAGCCCGTCTACGCCCGC 162
```

```
Qy    222 GATGCCAACTTCCAGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCG 281
Db    163 AACGCAAACTTCCAGCGCATCACCGACCCCAACGCCAAGTCCGGCTGCGATGGCGGCTCC 222

Qy    282 GCCTACTCCTGCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTC 341
Db    223 GCCTTCTCCTGCGCCGACCAGACCCCGTGGGCCGTGAGCGACGACTTTGCCTACGGTTTC 282

Qy    342 GCCGCGACGAGCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTC 401
Db    283 GCGGCTACGGCGCTCGCCGGCCAGTCCGAGTCTTCGTGGTGCTGTGCCTGCTACGAACTC 342

Qy    402 ACCTTCACTTCCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGC 461
Db    343 ACCTTCACTTCGGGCCCCGTTGCTGGCAAGAAGATGGCTGTCCAGTCCACCAGCACTGGC 402

Qy    462 GGCGACCTGGGAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTC 521
Db    403 GGTGACCTCGGTAGCAACCACTTTGACCTCAACATGCCAGGTGGCGGTGTCGGCATCTTC 462

Qy    522 AACGGCTGCAGCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCG 581
Db    463 GACGGCTGCTCGCCTCAGGTTGGCGGTCTCGCCGGCCAGCGCTATGGCGGCGTCTCGTCC 522

Qy    582 CGCGACCAGTGCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGAC 641
Db    523 CGCAGCGAATGCGACTCCTTCCCCGCGGCACTCAAGCCCGGCTGCTACTGGCGCTACGAC 582

Qy    642 TGGTTCCAGAACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAG 701
Db    583 TGGTTTAAGAACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAG 642

Qy    702 ATCGTTGCCCGCTCCGGCTGCAAGCGGAACGACGACTCCAGCTTCCCCGTCTTCACCCCC 761
Db    643 CTCGTCGCTCGCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATC 702

Qy    762 CCAAGCGGTGGCAACGG----------------TGGCACCGGGACGCCCACGTCGACT 803
Db    703 CCCTCCAGCAGCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACC 762

Qy    804 GCGCCTGGGTCGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAG 860
Db    763 TCCACCACCTCGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGG 822

Qy    861 TGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACC 920
Db    823 TGGGCTCAGTGCGGCGGCAATGGCTGGAGCGGGTGCACCACCTGCGTCGCTGGCAGCACT 882

Qy    921 TGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960
Db    883 TGCACGAAGATTAATGACTGGTACCATCAGTGCCTGTAGA 922
```

```
RESULT 6
A21793
LOCUS       A21793                 1060 bp    mRNA    linear   PAT 20-SEP-1995
DEFINITION  H.insolens mRNA for endonuclease component.
ACCESSION   A21793
VERSION     A21793.1  GI:1246872
KEYWORDS    .
SOURCE      Humicola insolens.
  ORGANISM  Humicola insolens
            Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
FEATURES             Location/Qualifiers
     source          1. .1060
                     /organism="Humicola insolens"
                     /strain="DSM 1800"
                     /db_xref="taxon:34413"
     CDS             10. .927
                     /codon_start=1
                     /product="endonuclease"
                     /protein_id="CAA01574.1"
                     /db_xref="GI:1246873"
                     /translation="MRSSPLLPSAVVAALPVLALAADGRSTRYWDCCKPSCGWAKKAP
                     VNQPVFSCNANFQRITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSN
                     EAGWCCACYELTFTSGPVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQF
                     GGLPGQRYGGISSRNECDRFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVART
                     GCRRNDDGNFPAVQIPSSSTSSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQ
                     CGGNGWSGCTTCVAGSTCTKINDWYHQCL"
BASE COUNT      190 a    377 c    288 g    205 t
ORIGIN
```

Query Match 37.3%; Score 437.4; DB 6; Length 1060;
Best Local Similarity 70.4%; Pred. No. 3.3e-67;
Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

```
Qy     55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
Db      5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy    115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
Db     65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy    175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
Db    119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy    235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
Db    179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy    292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
Db    239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy    352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
Db    299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy    412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
Db    359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy    472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
Db    419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy    532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
Db    479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy    592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
Db    539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy    652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy    772 GCAACGG----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
Db    719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy    814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898
```

```
Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
          | || ||||  |||   |||||||| || ||
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 7
A23635
LOCUS       A23635                  1060 bp    mRNA    linear   PAT 23-FEB-1995
DEFINITION  H. insolens endoglucanase gene.
ACCESSION   A23635
VERSION     A23635.1  GI:832886
KEYWORDS    .
SOURCE      Humicola insolens.
  ORGANISM  Humicola insolens
            Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
REFERENCE   1  (bases 1 to 1060)
  AUTHORS   Convents,A.C., Busch,A. and Baeck,A.C.
  TITLE     Detergent compositions with high activity cellulase and softening
            clays
  JOURNAL   Patent: EP 0495258-A 1 22-JUL-1992;
            THE PROCTER & GAMBLE COMPANY
FEATURES             Location/Qualifiers
     source          1. .1060
                     /organism="Humicola insolens"
                     /db_xref="taxon:34413"
     CDS             10. .927
                     /codon_start=1
                     /product="endoglucanase"
                     /protein_id="CAA01681.1"
                     /db_xref="GI:832887"
                     /translation="MRSSPLLPSAVVAALPVLALAADGRSTRYWDCCKPSCGWAKKAP
                     VNQPVFSCNANFQRITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSN
                     EAGWCCACYELTFTSGPVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQF
                     GGLPGQRYGGISSRNECDRFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVART
                     GCRRNDDGNFPAVQIPSSSTSSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQ
                     CGGNGWSGCTTCVAGSTCTKINDWYHQCL"
     sig_peptide     10. .72
BASE COUNT      190 a    377 c    288 g    205 t
ORIGIN

  Query Match             37.3%;  Score 437.4;  DB 6;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3.3e-67;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy     55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
          ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db      5 CCAAGATGCGTTCCTCCCCCCTCCTCCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy    115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
          ||   || ||   ||||||      |||||| | ||||||||||||||||||| ||||
Db     65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy    175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
          |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db    119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy    235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
          ||||  |  | |||||| | | | |||| |||||| |    |||||    ||||||||
Db    179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy    292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
          |||||||||||| || ||||| |||||||| |  ||||   ||| || || || ||
Db    239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy    352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
            || |||||| |    ||  |    ||||||||||||||||||| ||||||||||| |
Db    299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy    412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
          ||||||| || || |||||||  ||||| || ||||| || |||||||||| || || |
Db    359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy    472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
          | || ||||| |||||| ||  ||| |||||||||||||| ||||||||| |||| ||||
Db    419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy    532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
          | |||||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db    479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy    592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
          |||||   |||||||   | ||||||||||||||| | ||||| || ||||||||| |||
Db    539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy    652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
          |||||||||| |||| |||| |||| ||||| |||||||| |||||| |||| || |
Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
          || |||| |||   |||||||||||  || |||||| | | ||    |||   | |
Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy    772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
          ||| | |                   |  || |   |  ||||||| ||  |  |    |
Db    719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy    814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
          || ||| | |       ||          ||| ||||||||  || ||| ||||||||||
Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
          |||| |||  |||| ||||| |||||||||| ||| |||||| ||| |||  ||||
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
          | || ||||  |||   |||||||| || ||
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 8
A23644
LOCUS       A23644                  1060 bp    mRNA    linear   PAT 23-FEB-1995
DEFINITION  H. insolens endoglucanase gene.
ACCESSION   A23644
VERSION     A23644.1  GI:832890
KEYWORDS    .
SOURCE      Humicola insolens.
  ORGANISM  Humicola insolens
            Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
REFERENCE   1  (bases 1 to 1060)
  AUTHORS   Convents,A.C., Busch,A. and Baeck,A.C.
  TITLE     Detergent compositions with high activity cellulase and quaternary
            ammonium compounds
  JOURNAL   Patent: EP 0495554-A 1 22-JUL-1992;
            THE PROCTER & GAMBLE COMPANY
FEATURES             Location/Qualifiers
     source          1. .1060
                     /organism="Humicola insolens"
                     /strain="DSM 1800"
                     /db_xref="taxon:34413"
     CDS             10. .927
                     /codon_start=1
                     /product="endoglucanase"
                     /protein_id="CAA01683.1"
                     /db_xref="GI:832891"
                     /translation="MRSSPLLPSAVVAALPVLALAADGRSTRYWDCCKPSCGWAKKAP
                     VNQPVFSCNANFQRITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSN
                     EAGWCCACYELTFTSGPVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQF
                     GGLPGQRYGGISSRNECDRFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVART
                     GCRRNDDGNFPAVQIPSSSTSSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQ
                     CGGNGWSGCTTCVAGSTCTKINDWYHQCL"
     sig_peptide     10. .72
BASE COUNT      190 a    377 c    288 g    205 t
ORIGIN

  Query Match             37.3%;  Score 437.4;  DB 6;  Length 1060;
```

Best Local Similarity 70.4%; Pred. No. 3.3e-67;
Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

```
Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           ||| ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | |||||| | | | |||| |||||| |    |||||    |||||||| |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||||| || ||||| |||||||| |  ||||   ||| || || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||| |    ||  |    ||||||||||||||||||| |||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           ||||||| || || |||||||  ||||| || ||||| || ||||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| ||||||||||||| |||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
             | |||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | ||||||||||||||| | ||||| || |||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| |||| |||| |||| ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||  |||||||||||  || ||||||  | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                 |  || |   |   ||||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||          ||| ||||||||  || ||| ||||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| |||| |||| ||||| |||||||||||| ||| |||||| ||| ||| ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || ||||  |||   |||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929
```

RESULT 9
A23953
LOCUS A23953 1060 bp DNA linear PAT 01-MAR-1995
DEFINITION Endoglucanase gene.
ACCESSION A23953

VERSION A23953.1 GI:832894
KEYWORDS .
SOURCE Humicola insolens.
ORGANISM Humicola insolens
Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
REFERENCE 1 (bases 1 to 1060)
AUTHORS Baeck,A.C., Busch,A. and Ceulemans,R.A.A.
TITLE Compact detergent compositions with high activity cellulase
JOURNAL Patent: EP 0495257-A 1 22-JUL-1992;
THE PROCTER & GAMBLE COMPANY
FEATURES Location/Qualifiers
source 1. .1060
/organism="Humicola insolens"
/db_xref="taxon:34413"
CDS 10. .927
/codon_start=1
/product="endoglucanase"
/protein_id="CAA01696.1"
/db_xref="GI:832895"
/translation="MRSSPLLPSAVVAALPVLALAADGRSTRYWDCCKPSCGWAKKAP
VNQPVFSCNANFQRITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSN
EAGWCCACYELTFTSGPVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQF
GGLPGQRYGGISSRNECDRFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVART
GCRRNDDGNFPAVQIPSSSTSSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQ
CGGNGWSGCTTCVAGSTCTKINDWYHQCL"
BASE COUNT 190 a 377 c 288 g 205 t
ORIGIN

Query Match 37.3%; Score 437.4; DB 6; Length 1060;
Best Local Similarity 70.4%; Pred. No. 3.3e-67;
Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

```
Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           ||| ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | |||||| | | | |||| |||||| |    |||||    |||||||| |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||||| || ||||| |||||||| |  ||||   ||| || || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||| |    ||  |    ||||||||||||||||||| |||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           ||||||| || || |||||||  ||||| || ||||| || ||||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| ||||||||||||| |||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
             | |||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | ||||||||||||||| | ||||| || |||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
```

```
Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy    772 GCAACGG----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
Db    719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy    814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929
```

RESULT 10
A23957
LOCUS A23957 1060 bp DNA linear PAT 01-MAR-1995
DEFINITION Endoglucanase gene.
ACCESSION A23957
VERSION A23957.1 GI:832898
KEYWORDS .
SOURCE Humicola insolens.
ORGANISM Humicola insolens
Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
REFERENCE 1 (bases 1 to 1060)
AUTHORS McCorquodale,F. and Busch,A.
TITLE Dye transfer inhibiting compositions
JOURNAL Patent: EP 0540784-A 1 12-MAY-1993;
THE PROCTER & GAMBLE COMPANY
FEATURES Location/Qualifiers
source 1. .1060
/organism="Humicola insolens"
/db_xref="taxon:34413"
CDS 10. .927
/codon_start=1
/product="endoglucanase"
/protein_id="CAA01698.1"
/db_xref="GI:832899"
/translation="MRSSPLLPSAVVAALPVLALAADGRSTRYWDCCKPSCGWAKKAP
VNQPVFSCNANFQRITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSN
EAGWCCACYELTFTSGPVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQF
GGLPGQRYGGISSRNECDRFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVART
GCRRNDDGNFPAVQIPSSSTSSPVNQPTSTSTTSTSPPVQPTTPSGCTAERWAQ
CGGNGWSGCTTCVAGSTCTKINDWYHQCL"
BASE COUNT 190 a 377 c 288 g 205 t
ORIGIN

Query Match 37.3%; Score 437.4; DB 6; Length 1060;
Best Local Similarity 70.4%; Pred. No. 3.3e-67;
Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

```
Qy     55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
Db      5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy    115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
Db     65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy    175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
Db    119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy    235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
Db    179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy    292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
Db    239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy    352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
Db    299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy    412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
Db    359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy    472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
Db    419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy    532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
Db    479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy    592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
Db    539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy    652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy    772 GCAACGG----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
Db    719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy    814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929
```

RESULT 11
A41658
LOCUS A41658 1060 bp DNA linear PAT 05-MAR-1997
DEFINITION Sequence 1 from Patent EP0633311.
ACCESSION A41658
VERSION A41658.1 GI:2297284
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1060)
AUTHORS Herbots,I.M. and Jansen,M.P.
TITLE Hydrophobic amines for cellulase stabilization in liquid detergent compositions containing anionic surfactant and cellulase
JOURNAL Patent: EP 0633311-A 1 11-JAN-1995;
PROCTER & GAMBLE (US)
COMMENT Other publication CA 2165771 950105.
FEATURES Location/Qualifiers
source 1. .1060
/organism="unidentified"

```
                     /db_xref="taxon:32644"
BASE COUNT      190 a    377 c    288 g    205 t
ORIGIN

  Query Match             37.3%;  Score 437.4;  DB 6;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3.3e-67;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | |||||| | | | |||| |||||| |    |||||    |||||||| |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           ||||||||||||| || ||||| |||||||||  |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||| |    ||  |    ||||||||||||||||||| ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           ||||||| || || |||||||  |||| || ||||| || |||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| ||||||||||||| ||||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
             | |||||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | ||||||||||||||| | |||| || |||||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| ||||  ||||  ||||  ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   ||||||||||||  || |||||| | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG------------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           |||  | |                  |  || |   |   |||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||        ||| ||||||||  || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| ||||  ||||  ||||| |||||||||||| ||| |||||| ||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || ||||  |||   |||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929
```

```
RESULT 12
A68060
LOCUS       A68060                 1060 bp    DNA     linear   PAT 05-MAY-1999
DEFINITION  Sequence 55 from Patent WO9743409.
ACCESSION   A68060
VERSION     A68060.1  GI:4756862
KEYWORDS    .
SOURCE      Humicola insolens.
  ORGANISM  Humicola insolens
            Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
REFERENCE   1  (bases 1 to 1060)
  AUTHORS   Dalboege,H., Diderichsen,B., Sandal,T. and Kauppinen,S.
  TITLE     METHOD OF PROVIDING NOVEL DNA SEQUENCES
  JOURNAL   Patent: WO 9743409-A 55 20-NOV-1997;
            NOVONORDISK AS (DK)
FEATURES             Location/Qualifiers
     source          1. .1060
                     /organism="Humicola insolens"
                     /strain="DSM 1800"
                     /db_xref="taxon:34413"
     CDS             10. .927
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAB42307.1"
                     /db_xref="GI:4756863"
                     /translation="MRSSPLLPSAVVAALPVLALAADGRSTRYWDCCKPSCGWAKKAP
                     VNQPVFSCNANFQRITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSN
                     EAGWCCACYELTFTSGPVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQF
                     GGLPGQRYGGISSRNECDRFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVART
                     GCRRNDDGNFPAVQIPSSSTSSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQ
                     CGGNGWSGCTTCVAGSTCTKINDWYHQCL"
     sig_peptide     10. .72
     mat_peptide     73. .924
                     /product="unnamed"
BASE COUNT      190 a    377 c    288 g    205 t
ORIGIN

  Query Match             37.3%;  Score 437.4;  DB 6;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3.3e-67;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | |||||| | | | |||| |||||| |    |||||    |||||||| |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           ||||||||||||| || ||||| |||||||||  |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||| |    ||  |    ||||||||||||||||||| ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           ||||||| || || |||||||  |||| || ||||| || |||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| ||||||||||||| ||||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478
```

```
Qy    532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
Db    479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy    592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
Db    539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy    652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy    772 GCAACGG----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
Db    719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy    814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929
```

```
RESULT 13
AR059002
LOCUS       AR059002               1060 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 7 from patent US 5837847.
ACCESSION   AR059002
VERSION     AR059002.1  GI:5984579
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1060)
  AUTHORS   Royer,J.C., Moyer,D.L., Wendy,Y.T. and Shuster,J.R.
  TITLE     Non-toxic, non-toxigenic, non-pathogenic fusarium expression system
            and promoters and terminators for use therein
  JOURNAL   Patent: US 5837847-A 7 17-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .1060
                     /organism="unknown"
BASE COUNT      190 a    377 c    288 g    205 t
ORIGIN

  Query Match             37.3%;  Score 437.4;  DB 6;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3.3e-67;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy     55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
Db      5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy    115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
Db     65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy    175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
Db    119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy    235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
Db    179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy    292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
Db    239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy    352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
Db    299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy    412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
Db    359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy    472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
Db    419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy    532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
Db    479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy    592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
Db    539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy    652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy    772 GCAACGG----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
Db    719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy    814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929
```

```
RESULT 14
AR072921
LOCUS       AR072921               1060 bp    DNA     linear   PAT 28-AUG-2000
DEFINITION  Sequence 1 from patent US 5948672.
ACCESSION   AR072921
VERSION     AR072921.1  GI:9999684
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1060)
  AUTHORS   Rasmussen,G., Mikkelsen,J.Moslashedller., Schulein,M.,
            Patkar,S.Anant., Hagen,F., Hjort,C.Mailand. and Hastrup,S.
  TITLE     Cellulase preparation comprising an endoglucanase enzyme
  JOURNAL   Patent: US 5948672-A 1 07-SEP-1999;
FEATURES             Location/Qualifiers
     source          1. .1060
                     /organism="unknown"
BASE COUNT      190 a    377 c    288 g    205 t
ORIGIN

  Query Match             37.3%;  Score 437.4;  DB 6;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 3.3e-67;
```

Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

```
Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      |||||| | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           ||| ||| ||  |||||| |||| | ||| || ||||   | ||| |.||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | |||||| | | | |||| |||||| |    |||||    |||||||| |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||||| || ||||| |||||||||| |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||| |    ||  |   |||||||||||||||||| ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           ||||||| || || ||||||| ||||| || ||||| || |||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| ||||||||||||| ||||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
             | |||||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | ||||||||||||||| | ||||| || |||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| |||| |||| |||| ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   |||||||||||  || |||||| | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                 |  || |   |  ||||||| ||  |  |   |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |      ||        ||| |||||||| || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| ||||  ||||  ||||| |||||||||| ||| |||||| ||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || ||||  |||   |||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929
```

RESULT 15
AR163160
LOCUS AR163160 1060 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 55 from patent US 6270968.
ACCESSION AR163160
VERSION AR163160.1 GI:16233662
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1060)
AUTHORS Dalb.o slashed.ge,H., Sandal,T., Kauppinen,M.Sakari. and Diderichsen,Bslashedrge.
TITLE Method of providing a hybrid polypeptide exhibiting an activity of interest
JOURNAL Patent: US 6270968-A 55 07-AUG-2001;
FEATURES Location/Qualifiers
source 1. .1060
/organism="unknown"
BASE COUNT 190 a 377 c 288 g 205 t
ORIGIN

Query Match 37.3%; Score 437.4; DB 6; Length 1060;
Best Local Similarity 70.4%; Pred. No. 3.3e-67;
Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

```
Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      |||||| | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           ||| ||| ||  |||||| |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | |||||| | | | |||| |||||| |    |||||    |||||||| |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||||| || ||||| |||||||||| |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||| |    ||  |   |||||||||||||||||| ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           ||||||| || || ||||||| ||||| || ||||| || |||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| ||||||||||||| ||||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
             | |||||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | ||||||||||||||| | ||||| || |||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| |||| |||| |||| ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   |||||||||||  || |||||| | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                 |  || |   |  ||||||| ||  |  |   |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
```

```
          || ||| | |       ||         ||| ||||||||  || ||| ||||||||||
Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
          |||| ||||  ||||  ||||| |||||||||| ||| |||||| ||| |||  ||||
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
          | || |||| |||  |||||||| || ||
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


Search completed: February 27, 2003, 19:35:59
Job time : 3189 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: February 27, 2003, 17:47:15 ; Search time 305 Seconds
(without alignments)
8668.348 Million cell updates/sec

Title: US-10-007-521-11
Perfect score: 1174
Sequence: 1 GAGCAGCACCCCTCAAGCTG.....AAAAAAAAAAAAAAAAAAAA 1174

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1174	100.0	1174	17	AAT39050	cDNA encoding cell
2	1174	100.0	1174	19	AAV39096	Monocomponent endo
3	453.6	38.6	1261	19	AAV23748	Humicola grisea ce
4	451.6	38.5	922	19	AAV15072	Hybrid DNA compris
5	439	37.4	1060	16	AAZ60178	H. insolens endogl
6	437.4	37.3	1058	13	AAQ26405	Humicola insolens
7	437.4	37.3	1060	12	AAQ14856	Humicola insolens
8	437.4	37.3	1060	13	AAQ25932	Cellulase containe
9	437.4	37.3	1060	13	AAQ29934	Endoglucanase gene
10	437.4	37.3	1060	13	AAQ30067	Sequence encoding
11	437.4	37.3	1060	14	AAQ41732	Dye transfer inhib
12	437.4	37.3	1060	14	AAQ49941	Endoglucanase enzy
13	437.4	37.3	1060	17	AAT10182	Alkaline endogluca
14	437.4	37.3	1060	19	AAV16102	Humicola insolens
15	437.4	37.3	1060	19	AAV15065	Humicola insolens
16	437.4	37.3	1060	21	AAA53334	Endoglucanase nucl
17	435.8	37.1	1060	13	AAQ26380	Endoglucanase #1.
18	431	36.7	1060	13	AAQ30072	43kD endoglucanase
19	429.4	36.6	1060	13	AAQ31181	H. insolens cellul
20	412.8	35.2	922	19	AAV15073	Hybrid DNA compris
21	393.4	33.5	925	19	AAV15076	Hybrid DNA compris
22	390.6	33.3	807	19	AAV16104	Humicola insolens
23	377.2	32.1	1257	19	AAV19378	Humicola insolens
24	377.2	32.1	1257	19	AAV13840	Humicola insolens
25	377.2	32.1	1257	24	ABL59235	Nucleotide sequenc
26	377.2	32.1	1257	24	AAL43251	Humicola insolens
27	377.2	32.1	1257	24	ABL60700	H. insolens EGV nu
28	368.2	31.4	1154	17	AAT39048	cDNA encoding cell
29	366.4	31.2	928	19	AAV15074	Hybrid DNA compris
30	365.6	31.1	927	19	AAV29596	Humicola insolens
31	353.6	30.1	936	18	AAT66541	20K-cellulase gene
32	337.6	28.8	915	19	AAV15075	Hybrid DNA compris
33	337	28.7	894	17	AAT39061	Chimeric endogluca
34	334.8	28.5	927	17	AAT39062	Chimeric endogluca
35	315.2	26.8	1473	12	AAQ14857	Fusarium oxysporum
36	315.2	26.8	1473	13	AAQ26407	Fusarium oxysporum
37	315.2	26.8	1473	13	AAQ26382	Endoglucanase #2.
38	315.2	26.8	1473	13	AAQ25933	Cellulase containe
39	315.2	26.8	1473	13	AAQ29935	Endoglucanase gene
40	315.2	26.8	1473	14	AAQ49942	Endoglucanase enzy
41	315.2	26.8	1473	16	AAZ60179	F. oxysporum endog
42	315.2	26.8	1473	19	AAV16103	Fusarium oxysporum
43	313.6	26.7	885	17	AAT39075	Chimeric endogluca
44	313.6	26.7	1473	14	AAQ41733	Dye transfer inhib
45	302	25.7	1132	17	AAT39053	cDNA encoding cell

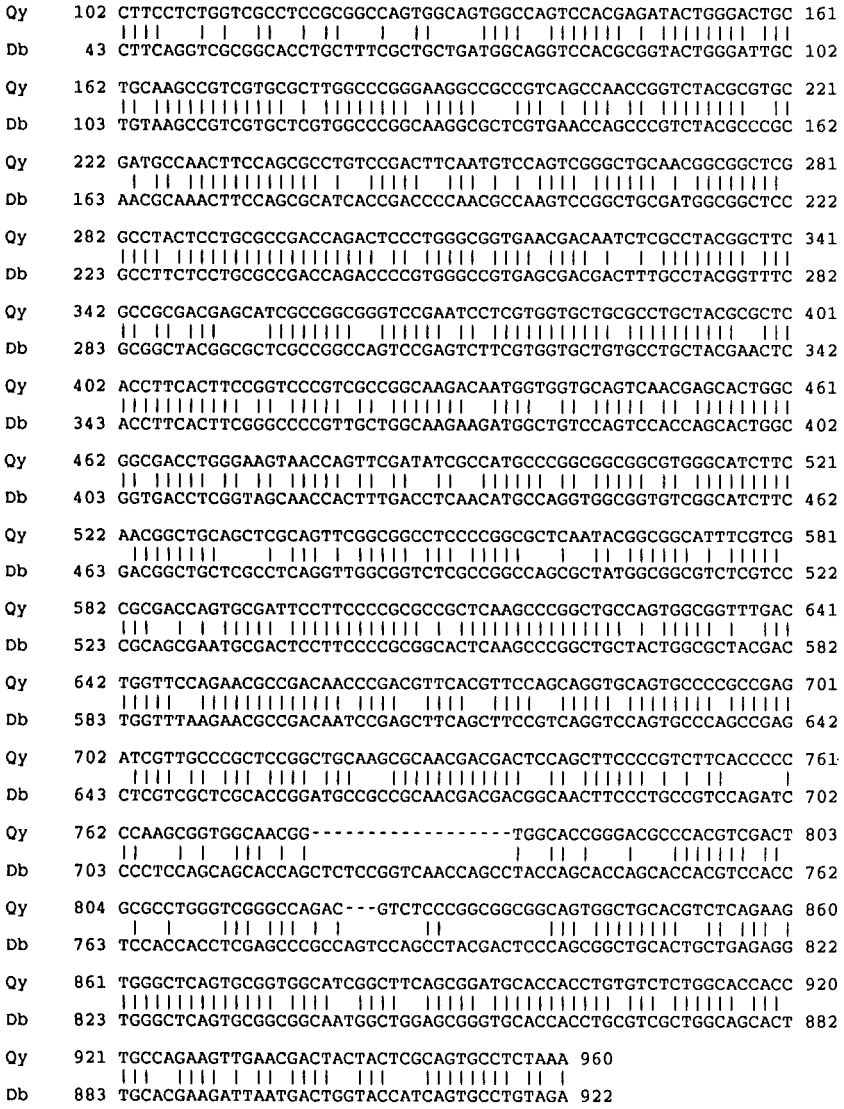
ALIGNMENTS

RESULT 1
AAT39050
ID AAT39050 standard; cDNA; 1174 BP.
XX
AC AAT39050;
XX
DT 20-MAY-1997 (first entry)
XX
DE cDNA encoding cellulytic enzyme #4 of the invention.
XX
KW Cellulytic enzyme; endoglucanase; hydrolysis; cellulose; microorganism;
KW plant cellulase; catalytic region; textile; backstaining; bio-polishing;
KW stone-washing; cellulosic fabric; colour clarification; defibration;
KW cell wall degradation; paper pulp; debarking; fibre modification;
KW enzymatic de-inking; drainage improvement; ss.
XX
OS Thielavia terrestris.
XX
FH Key Location/Qualifiers
FT CDS 59..959
FT /*tag= a
FT /product= endoglucanase
XX
PN WO9629397-A1.
XX
PD 26-SEP-1996.
XX
PF 18-MAR-1996; 96WO-DK00105.
XX
PR 12-FEB-1996; 96DK-0000137.
PR 17-MAR-1995; 95DK-0000272.
PR 08-AUG-1995; 95DK-0000885.

```
PR   08-AUG-1995;   95DK-0000886.
PR   08-AUG-1995;   95DK-0000887.
PR   08-AUG-1995;   95DK-0000888.
XX
PA   (NOVO ) NOVO-NORDISK AS.
XX
PI   Andersen LN,  Ihara M,  Kauppinen MS,  Lange L,  Lassen SF;
PI   Nielsen RI, Schuelein M,  Takagi S;
XX
DR   WPI; 1996-443173/44.
DR   P-PSDB; AAW04928.
XX
PT   New endo:glucanase enzyme preparations - contg. conserved catalytic
PT   regions, useful for treating fabrics, textiles, plant material or
PT   paper pulp
XX
PS   Claim 72; Page 160-161; 316pp; English.
XX
CC   AAT39047-T39054 represent the coding sequences for the enzymes of the
CC   invention.  The encoded enzymes possess cellulytic (particularly
CC   endoglucanase) activity.  Cellulytic enzymes are involved in the the
CC   hydrolysis of cellulose, and are synthesised by a large number of
CC   microorganisms and plants.  The enzymes of the invention containing the
CC   conserved catalytic regions (such as AAW04913) exhibit improved
CC   performance, e.g. 50 times higher performance, compared to multiple
CC   domain enzymes.  The enzymes can be used for the treatment of fabrics or
CC   textiles, preferably for preventing backstaining, for bio-polishing or
CC   for stone-washing cellulosic fabric. They can also be used to provide
CC   colour clarification for laundry. The enzymes can also be used for the
CC   degradation or modification of plant material, such as cell walls.  They
CC   can also be used in the treatment of paper pulp preferably for
CC   debarking, defibration, fibre modification, enzymatic de-inking or
CC   drainage improvement.
XX
SQ   Sequence 1174 BP; 243 A; 395 C; 320 G; 216 T; 0 other;

  Query Match           100.0%;  Score 1174;  DB 17;  Length 1174;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-203;
  Matches 1174;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60

Qy      61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120

Qy     121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180

Qy     181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240

Qy     241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300

Qy     301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360

Qy     361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420

Qy     421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480

Qy     481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540

Qy     541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600

Qy     601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660

Qy     661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720

Qy     721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780

Qy     781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840

Qy     841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900

Qy     901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960

Qy     961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020

Qy    1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080

Qy    1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAA 1140

Qy    1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
          ||||||||||||||||||||||||||||||||||
Db    1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174


RESULT 2
AAV39096
ID   AAV39096 standard; cDNA; 1174 BP.
XX
AC   AAV39096;
XX
DT   21-SEP-1998  (first entry)
XX
DE   Monocomponent endoglucanase encoding cDNA.
XX
KW   Monocomponent endoglucanase; cellulolytic enzyme; garment; abrasion;
KW   abraded looking jeans; fungus; Thielavia terrestris; stone-washed; ss.
XX
OS   Thielavia terrestris.
XX
FH   Key             Location/Qualifiers
FT   CDS             60..959
FT                   /*tag=  a
FT                   /product= "monocomponent endoglucanase"
XX
PN   EP843041-A1.
XX
PD   20-MAY-1998.
XX
PF   30-MAY-1997;   97EP-0610021.
```

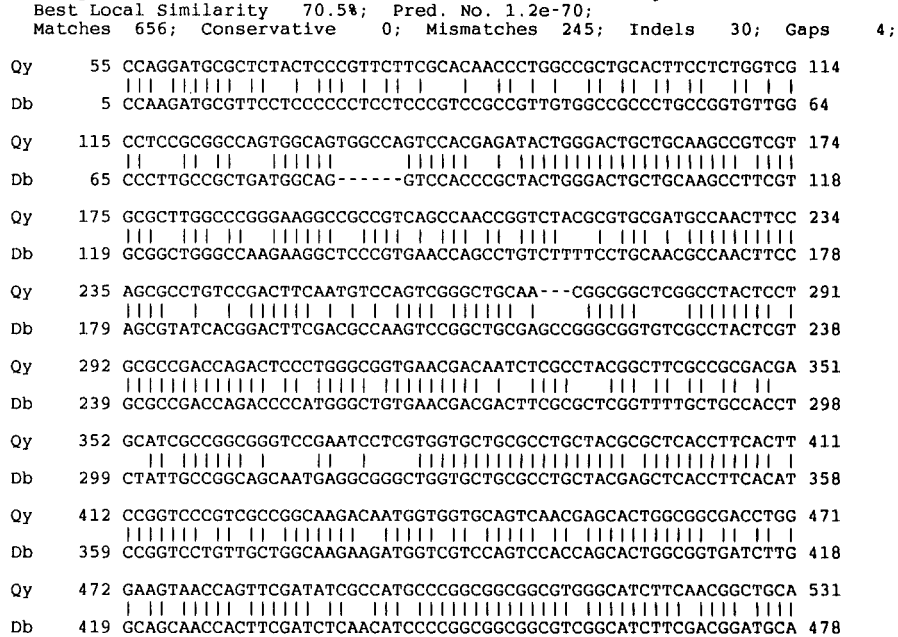
```
XX
PR   13-NOV-1996;   96DK-0001276.
XX
PA   (NOVO ) NOVO-NORDISK AS.
XX
PI   Kalum L,  Lund H;
XX
DR   WPI; 1998-263308/24.
DR   P-PSDB; AAW63624.
XX
PT   Pair of jeans - with abraded look
XX
PS   Claim 17; Pages 9-10; 16pp; English.
XX
CC   This cDNA encodes a monocomponent endoglucanase. This is a cellulolytic
CC   enzyme belonging to the fungal species Thielavia terrestris. This can be
CC   used in the process of invention of providing a pair of jeans made from
CC   dyed twill fabric that has localised variations in the colour density of
CC   the fabric. The process involves providing the jeans with a stone-washed
CC   or abraded look, where the variation corresponds to a delta remission
CC   value (Delta R) higher than 11, and the reflection of a first area of
CC   the jeans fabric is less than 12 percent. The process comprises selecting
CC   the desired textile cutting pattern for jeans garment, positioning the
CC   pattern onto newly manufactured dyed twill fabric, cutting the jeans
CC   garment parts, sowing the pair of jeans, optionally subjecting the pair
CC   of jeans to a desizing treatment, and subjecting the pair of jeans to an
CC   abrasion treatment with an efficient amount of a cellulolytic enzyme in
CC   an aqueous medium essentially free of bleaching chemicals. The abraded
CC   look near stitching is different from the abraded look far from
CC   stitching.
XX
SQ   Sequence 1174 BP; 243 A; 395 C; 320 G; 216 T; 0 other;

  Query Match             100.0%;  Score 1174;  DB 19;  Length 1174;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-203;
  Matches 1174;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAGCAGCACCCCTCAAGCTGTACAGTTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGA 60

Qy      61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCG 120

Qy     121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTT 180

Qy     181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCC 240

Qy     241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGTCCGACTTCAATGTCCAGTCGGGCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACC 300

Qy     301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCG 360

Qy     361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCG 420

Qy     421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACC 480

Qy     481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGT 540

Qy     541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCT 600

Qy     601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACA 660

Qy     661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCT 720

Qy     721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTG 780

Qy     781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GCACCGGGACGCCCACGTCGACTGCGCCTGGGTCGGGCCAGACGTCTCCCGGCGGCGGCA 840

Qy     841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTGGCATCGGCTTCAGCGGATGCACCA 900

Qy     901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACGACTACTACTCGCAGTGCCTCTAAA 960

Qy     961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CAGCTTTTCGCACGAGGTGGCGGGACGGAGCAAGGAGACCGTCAACTTCGTCATGCATAT 1020

Qy    1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACA 1080

Qy    1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAA 1140

Qy    1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
          ||||||||||||||||||||||||||||||||||
Db    1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174


RESULT 3
AAV23748
ID   AAV23748 standard; cDNA; 1261 BP.
XX
AC   AAV23748;
XX
DT   19-AUG-1998  (first entry)
XX
DE   Humicola grisea cellulase cDNA sequence.
XX
KW   Cellulase; variant; cellulolytic enzyme; endoglucanase V; EGV; paper;
KW   protein engineering; detergent; fabric softener; pulp; debarking;
KW   defibration; fibre modification; degradation; ss.
XX
OS   Humicola grisea.
XX
PN   WO9812307-A1.
XX
PD   26-MAR-1998.
XX
PF   17-SEP-1997;   97WO-DK00393.
XX
PR   17-SEP-1996;   96DK-0001013.
XX
PA   (NOVO ) NOVO-NORDISK AS.
XX
PI   Andersen KV,  Christiansen L,  Damgaard B,  Schulein M;
XX
```

```
DR   WPI; 1998-217251/19.
XX
PT   Cellulase enzyme variants - having amino acid changes which improve
PT   properties e.g. activity, sensitivity to surfactants, pH optimum or
PT   stability
XX
PS   Example 3; Page 77-78; 115pp; English.
XX
CC   The present invention describes a cellulase enzyme variant comprising a
CC   catalytic core domain exhibiting cellulolytic activity which is derived
CC   from a naturally occurring parental cellulase by amino acid residue
CC   substitution, insertion and/or deletion; and (with cellulase numbering):
CC   (a) at position 5 has an Ala, Ser, or Thr residue; (b) at position 8 has
CC   a Phe or a Tyr residue; (c) at position 9 has a Phe, Trp, or Tyr residue;
CC   (d) at position 10 has a Asp residue; (e) at position 121 has a Asp
CC   residue. The present sequence represents the cDNA sequence of Humicola
CC   grisea cellulase from the present invention. The new cellulase enzymes
CC   can be used in e.g. detergent or fabric softener compositions, for bio-
CC   polishing of new fabrics, for promoting a stone-washed look to cellulosic
CC   containing fabric, for pulp and paper applications, e.g. for debarking,
CC   defibration, fibre modification, drainage improvement, inter fibre
CC   bonding or for degradation of plant material e.g. for improving feed
CC   value. The cellulase variants have improved properties with respect to
CC   e.g. catalytic activity, altered sensitivity to anionic tensides, pH
CC   optimum or activity profile or stability.
XX
SQ   Sequence 1261 BP; 246 A; 417 C; 336 G; 262 T; 0 other;

  Query Match             38.6%;  Score 453.6;  DB 19;  Length 1261;
  Best Local Similarity   74.9%;  Pred. No. 2.8e-73;
  Matches  583;  Conservative    0;  Mismatches  189;  Indels    6;  Gaps    1;

Qy      26 TTTCCACCCCGCTCTCTTTTCTTCGGCCCCCAGGATGCGCTCTACTCCCGTTCTTCGCAC 85
           |||    | |||||  || |  |    |  ||  |||||||||| ||||  || | |||||
Db      18 TTTTATTCACGCTCATTTATTCTAAAACTTCAATATGCGCTCTGCTCCTATTTTCCGCAC 77

Qy      86 AACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCGCGGCCAGTGGCAGTGGCCAGTCCAC 145
           ||||||| ||||| || || || | |||   ||| |||  |||||      |||| ||
Db      78 GGCCCTGGCGGCTGCGCTCCCCCTTGCCGCACTCGCCGCCGATGGCA------AGTCGAC 131

Qy     146 GAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTTGGCCCGGGAAGGCCGCCGTCAGCCA 205
           |||||||||||||||||||||| ||||||  |||||||||  |||||   ||| | |||
Db     132 CAGATACTGGGACTGCTGCAAGCCATCGTGCTCTTGGCCCGGAAAGGCACTCGTGAACCA 191

Qy     206 ACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCCTGTCCGACTTCAATGTCCAGTCGGG 265
           || |||| | | ||||| |||||  |||||||||  |  |||||  ||||  | |||||||
Db     192 GCCTGTCTTCACTTGCGACGCCAAATTCCAGCGCATCACCGACCCCAATACCAAGTCGGG 251

Qy     266 CTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACCAGACTCCCTGGGCGGTGAACGACAA 325
           |||| | ||||||||||||||  || || || ||||||||  ||||||||  |||||||| |
Db     252 CTGCGATGGCGGCTCGGCCTTTTCGTGTGCTGACCAGACCCCCTGGGCTCTGAACGACGA 311

Qy     326 TCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTG 385
           | ||||||| |||||||| || |||   ||  | || || || ||| ||||||||||||
Db     312 TGTCGCCTATGGCTTCGCTGCCACGGCTATTTCGGGTGGATCGGAAGCCTCGTGGTGCTG 371

Qy     386 CGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCA 445
           ||| |||||||| || || ||||| || || || || ||||||||||| ||||| || ||
Db     372 CGCATGCTACGCTCTTACTTTCACCTCGGGCCCTGTGGCCGGCAAGACCATGGTCGTCCA 431

Qy     446 GTCAACGAGCACTGGCGGCGACCTGGGAAGTAACCAGTTCGATATCGCCATGCCCGGCGG 505
           ||| || | ||| |||||||| || || || ||||| |||||   ||   || || |||||
Db     432 GTCGACCAACACCGGCGGCGATCTCGGCAGCAACCATTTCGACCTCCAGATTCCAGGCGG 491

Qy     506 CGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATA 565
           ||| || ||||||||  | || |||| | | |||||||| || ||| | ||||  |  ||
Db     492 CGGTGTCGGCATCTTTGATGGGTGCACCCCCCAGTTCGGAGGTCTCGCTGGCGAACGCTA 551

Qy     566 CGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTG 625
           ||| ||||| ||    |||  |   ||||| || ||||| ||| ||||||||||||||||
Db     552 CGGTGGCATCTCAGACCGCAGCTCCTGCGACTCGTTCCCTGCGGCGCTCAAGCCCGGCTG 611

Qy     626 CCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACAACCCGACGTTCACGTTCCAGCAGGT 685
           ||  ||||| || || |||||| |||||||||||||||||| || || ||| |||||||
Db     612 CCTCTGGCGCTTCGATTGGTTCAAGAACGCCGACAACCCGACCTTTACCTTCAAGCAGGT 671

Qy     686 GCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCTGCAAGCGCAACGACGACTCCAGCTT 745
           ||||||||||||||| | |||||| |  |||||||||||| | || |||  || |||
Db     672 GCAGTGCCCCGCCGAGCTTGTTGCCAGGACCGGCTGCAAGCGCGAGGATGACGGCAACTT 731

Qy     746 CCCCGTCTTCACCCCCCCAAGCGGTGGCAACGGTGGCACCGGGACGCCCACGTCGACT 803
           |||||||||||| |||||    ||| |||||   |||     |   | | |  || |||
Db     732 CCCCGTCTTCACGCCCCCCGCGGGTAGCAACACCGGCGGTAGCCAGTCGAGCTCCACT 789


RESULT 4
AAV15072
ID   AAV15072 standard; DNA; 922 BP.
XX
AC   AAV15072;
XX
DT   19-MAY-1998  (first entry)
XX
DE   Hybrid DNA comprising a family 45 cellulase core region.
XX
KW   Cellulase core region; isolation; microorganism; identification;
KW   hybrid DNA; ss.
XX
OS   Chimeric - Humicola nigrescens.
OS   Chimeric - Humicola insolens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..921
FT                   /*tag=  a
FT                   /product= "hybrid gene construct"
XX
PN   WO9743409-A2.
XX
PD   20-NOV-1997.
XX
PF   12-MAY-1997;   97WO-DK00216.
XX
PR   10-MAY-1996;   96DK-0000562.
XX
PA   (NOVO ) NOVO-NORDISK AS.
XX
PI   Dalboge H,  Diderichsen B,  Kauppinen S,  Sandal T;
XX
DR   WPI; 1998-008878/01.
DR   P-PSDB; AAW44269.
XX
PT   Isolating novel DNA sequences from microorganisms - without the need
PT   for culturing the microorganism
XX
PS   Example 3; Page 51-52; 72pp; English.
XX
CC   The present sequence represents a novel hybrid gene construct
CC   from an example of the present invention. The present invention
CC   describes a novel method for providing a novel DNA sequence encoding a
CC   polypeptide from a microorganism with an activity of interest. The
CC   method comprises: (i) PCR amplification of the DNA with PCR primers with
CC   homology to (a) known gene(s) encoding a polypeptide with an activity of
CC   interest; (ii) linking the obtained PCR product of a 5' structural gene
CC   sequence and a 3' structural gene sequence; (iii) expressing the
CC   resulting hybrid DNA sequence; (iv) screening for hybrid DNA sequences
CC   encoding a polypeptide with the activity of interest or a related
CC   activity; and (v) isolating the hybrid DNA sequence identified in step
CC   (iv). This method provides for identification and isolation of sequences
CC   from microorganisms without having to cultivate and isolate the
CC   microorganism.
XX
SQ   Sequence 922 BP; 151 A; 333 C; 262 G; 176 T; 0 other;

  Query Match             38.5%;  Score 451.6;  DB 19;  Length 922;
  Best Local Similarity   71.6%;  Pred. No. 6.2e-73;
  Matches  630;  Conservative    0;  Mismatches  229;  Indels   21;  Gaps    2;
```



RESULT 5
AAZ60178
ID AAZ60178 standard; cDNA; 1060 BP.
XX
AC AAZ60178;
XX
DT 25-APR-2000 (first entry)
XX
DE H. insolens endoglucanase enzyme coding sequence.
XX
KW Endoglucanase; cellulase enzyme; detergent composition; anion surfactant;
KW degradation protection; liquid detergent; long term stability; ss.
XX
OS Humicola insolens.
XX
PN EP633311-A1.
XX
PD 11-JAN-1995.
XX
PF 28-JUN-1993; 93EP-0870122.
XX
PR 28-JUN-1993; 93EP-0870122.
XX
PA (PROC) PROCTER & GAMBLE CO.
XX
PI Herbots IMAJ, Jansen MP;
XX
DR WPI; 1995-038508/06.
DR P-PSDB; AAY67388.
XX
PT Liq. detergent compsns. - contg. hydrophobic amine(s) for cellulase
PT stabilisation
XX
PS Disclosure; Page 16-17; 28pp; English.
XX
CC This sequence represents the Humicola insolens endoglucanase enzyme
CC coding sequence. The enzyme is a cellulase enzyme used in the liquid
CC detergent composition of the invention. The detergent comprises an anion
CC surfactant, a cellulase enzyme and a cellulase stabilising amount of an
CC amine of the formula R1R2R3N; where R1 and R2 are independently H or a
CC C2-C9 alkyl chain, and R3 is a C2-C9 alkyl chain or cyclohexyl or
CC cyclopentyl or cycloheptyl. The amine prevents the cellulase enzyme from
CC becoming degraded. The composition is used as a liquid detergent, and has
CC long term stability, it is applicable to the protection of an cellulase
CC and can be used in the presence of proteases.
XX
SQ Sequence 1060 BP; 190 A; 378 C; 287 G; 205 T; 0 other;

Query Match 37.4%; Score 439; DB 16; Length 1060;
Best Local Similarity 70.5%; Pred. No. 1.2e-70;
Matches 656; Conservative 0; Mismatches 245; Indels 30; Gaps 4;



```
Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
           | ||||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | ||||||||||||||| | ||||| || ||||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| |||| |||| |||| ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   ||||||||||  || |||||| | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG------------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                  |  || |   |   ||||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||          ||| |||||||| || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| |||| |||| ||||| |||||||||||| ||| ||||||||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCACCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || ||||  |||  |||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 6
AAQ26405
ID   AAQ26405 standard; cDNA; 1058 BP.
XX
AC   AAQ26405;
XX
DT   11-JAN-1993  (first entry)
XX
DE   Humicola insolens DSM 1800 endoglucanase.
XX
KW   Detergent; washing powder; cellulase; softening clay; C14CMC-method;
KW   flocculation; radio-labelled carboxymethylcellulase; ss.
XX
OS   Humicola insolens DSM 1800.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     10..72
FT                   /*tag=  a
FT   mat_peptide     73..927
FT                   /*tag=  b
XX
PN   EP495258-A.
XX
PD   22-JUL-1992.
XX
PF   06-NOV-1991;   91EP-0202880.
XX
PR   16-JAN-1991;   91EP-0870006.
PR   06-NOV-1991;   91EP-0202880.
PR   06-NOV-1991;   91EP-0202879.
XX
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Baeck AC,  Busch A,  Convents AC;
XX
DR   WPI; 1992-243163/30.
DR   P-PSDB; AAR25525.
XX
PT   Compact, granular detergent compsns. - contain high activity
PT   cellulase and softening clay to provide synergistic effect in
PT   softening performance
XX
PS   Claim 10; Page 27-28; 38pp; English.
XX
CC   The endoglucanase isolated from H.insolens DSM 1800 has Mr 43,000
CC   and high cellulase activity (i.e. removes a minimum of 10% of
CC   radiolabelled carboxymethylcellulose (C14CMC)). The enzyme can be
CC   produced recombinantly using the cDNA coding sequence and used in
CC   detergent compositions along with a surface active agent, a builder
CC   system and a softening clay. See also AAQ26407.
XX
SQ   Sequence 1058 BP; 189 A; 376 C; 288 G; 205 T; 0 other;

  Query Match             37.3%;  Score 437.4;  DB 13;  Length 1058;
  Best Local Similarity   70.4%;  Pred. No. 2.4e-70;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||       ||||||  | ||||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | ||||||  | | | |||| ||||||  |    |||||    ||||||| |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||| || ||||| |||||||| |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||| |    ||  |    ||||||||||||||||||| |||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           ||||||| || || |||||||  ||||| || ||||| || |||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| |||||||||||||| |||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
             | ||||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | ||||||||||||||| | ||||| || |||||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| |||| |||| |||| ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   ||||||||||  || |||||| | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG------------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                  |  || |   |   ||||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||          ||| |||||||| || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838
```

```
Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
          |||| |||| |||| ||||| |||||||||||| ||| |||||| ||| |||  ||||
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
          | || |||| |||   |||||||| || ||
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 7
AAQ14856
ID   AAQ14856 standard; DNA; 1060 BP.
XX
AC   AAQ14856;
XX
DT   18-FEB-1992  (first entry)
XX
DE   Humicola insolens DSM 1800 endoglucanase.
XX
KW   cellulase; ss.
XX
OS   Humicola insolens.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     10..72
FT                   /*tag=  a
FT   mat_peptide     73..924
FT                   /*tag=  b
XX
PN   WO9117243-A.
XX
PD   14-NOV-1991.
XX
PF   08-MAY-1991;   91WO-DK00123.
XX
PR   22-APR-1991;   91DK-0000736.
PR   09-MAY-1990;   90DK-0001159.
XX
PA   (NOVO ) NOVO NORDISK A/S.
XX
PI   Rasmussen G,  Mikkelsen JM,  Schulein M,  Patkar SA,  Hagen F;
PI   Hjort CM,  Hastrup S;
XX
DR   WPI; 1991-353765/48.
DR   P-PSDB; AAR15271.
XX
PT   Cellulase prepn. comprising endoglucanase enzyme - used in
PT   detergents for cellulose-contg. fabrics or to improve drainage of
PT   paper pulp
XX
PS   Claim 11; Page 48; 67pp; English.
XX
CC   The cellulase (i.e. endoglucanase) gene was isolated from a H.
CC   insolens cDNA library by screening with a probe based on the amino
CC   acid sequence of the purified enzyme.
XX
SQ   Sequence 1060 BP; 190 A; 377 C; 288 G; 205 T; 0 other;

  Query Match             37.3%;  Score 437.4;  DB 12;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 2.4e-70;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps   4;

Qy     55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
          ||| |||||| ||  | ||| | || |    |  || | |  || || || ||  || | |
Db      5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy    115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
          ||   || ||   ||||||       ||||||  | ||||||||||||||||||||| ||||
Db     65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy    175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
          |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db    119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy    235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
          ||||  |  | |||||| | | | |||| |||||| |    |||||    |||||||| |
Db    179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy    292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
          |||||||||||||| || ||||| |||||||| |  ||||   ||| || || || ||
Db    239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy    352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
          || |||||| |    ||  |    |||||||||||||||||| ||||||||||| |
Db    299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy    412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
          ||||||| || || |||||||  ||||| || ||||| || |||||||||||| || || |
Db    359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy    472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
          | || ||||| |||||| ||  ||| |||||||||||||| |||||||| |||| ||||
Db    419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy    532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
          | ||||||||||| || ||||||    |  |||||||||| ||||| ||| || |||
Db    479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy    592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
          |||||   |||||||   | |||||||||||||| | ||||| || |||||||| |||
Db    539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy    652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
          |||||||||| |||| |||| |||| ||||| |||||||| |||||| |||| || |
Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
          || |||| |||   |||||||||||  || |||||| | | ||     |||   | |
Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy    772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
          ||| | |                   |  || |   |   ||||||| ||  |   |   |
Db    719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy    814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
          || ||| | |       ||         ||| ||||||||  || ||| ||||||||||
Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
          |||| |||| |||| |||||||||||| ||| |||||| ||| |||  ||||
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
          | || |||| |||   |||||||| || ||
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 8
AAQ25932
ID   AAQ25932 standard; DNA; 1060 BP.
XX
AC   AAQ25932;
XX
DT   15-JAN-1993  (first entry)
XX
DE   Cellulase contained in a detergent compsn.
XX
KW   Endoglucanase; immunoreactive; Humicola isolens; cleaning;
KW   softening; ss.
XX
OS   Humicola isolens.
XX
FH   Key             Location/Qualifiers
FT   CDS             9..927
FT                   /*tag=  a
XX
```

PN EP495554-A.
XX
PD 22-JUL-1992.
XX
PF 15-JAN-1992; 92EP-0200101.
XX
PR 16-JAN-1991; 91EP-0870006.
PR 06-NOV-1991; 91EP-0202879.
PR 06-NOV-1991; 91EP-0202881.
XX
PA (PROC) PROCTER & GAMBLE CO.
XX
PI Baeck AC, Boutique J, Bush A, Convents AC, Hargrove RS;
PI Prettyaj, Busch A, Ceulemans RAA;
XX
DR WPI; 1992-243405/30.
DR P-PSDB; AAR25428.
XX
PT Detergent compsn. for improving cleaning and performance - is
PT composed of quat. ammonium cpd. and high activity cellulose e.g.
PT homogeneous endo:glucanase
XX
PS Claim 10; Page 19; 28pp; English.
XX
CC The cellulase DNA was used in a detergent compsn. contg. a
CC quaternary ammonium cpd. The compsn. gives a detergent with good
CC cleaning and softening performance due to the synergistic effect of
CC the ammonium and cellolase components. The cellulase is a
CC homogeneous endoglucanase componant which is immunoreactive with an
CC antibody raised against a cellulase purified from Humicola isolens
CC DSM 1800. See also AAQ25933.
XX
SQ Sequence 1060 BP; 190 A; 377 C; 288 G; 205 T; 0 other;

Query Match 37.3%; Score 437.4; DB 13; Length 1060;
Best Local Similarity 70.4%; Pred. No. 2.4e-70;
Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

```
Qy     55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
          ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db      5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy    115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
          ||   || ||   ||||||      ||||||  | ||||||||||||||||||| ||||
Db     65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy    175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
          |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db    119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy    235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
          ||||  |  | ||||||  | | | |||| ||||||  |    |||||    |||||||  |
Db    179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy    292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
          |||||||||||| || ||||| |||||||||  |  ||||   ||| || || || ||
Db    239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy    352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
            || |||||| |    ||  |    ||||||||||||||||||| ||||||||||| |
Db    299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy    412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
          |||||||  || || |||||||  ||||| || ||||| || ||||||||||| || || |
Db    359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy    472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
          | || ||||| |||||| ||  ||| |||||||||||||| |||||||| |||| ||||
Db    419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy    532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
            | |||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db    479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy    592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
          |||||   |||||||   | |||||||||||||| | ||||| || ||||||||| |||
Db    539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy    652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
          |||||||||| |||| |||| |||| ||||| |||||||| |||||| |||| || | |
Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
          || |||| |||   |||||||||||  || |||||| | | ||     |||   | |
Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy    772 GCAACGG----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
          ||| | |                  |  || |   |   ||||||| ||  |  |    |
Db    719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy    814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
          || ||| | |       ||         ||| ||||||||  || ||| ||||||||||
Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
          |||| |||| |||| ||||| |||||||||||| ||| |||||| ||| |||  ||||
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
          | || ||||  |||   |||||||| || ||
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929
```

RESULT 9
AAQ29934
ID AAQ29934 standard; cDNA; 1060 BP.
XX
AC AAQ29934;
XX
DT 09-MAR-1993 (first entry)
XX
DE Endoglucanase gene.
XX
KW Alkaline cellulase; laundry detergent compositions;
KW fabric colour maintenance; ss.
XX
OS Humicola insolens DSM 1800.
XX

FH	Key	Location/Qualifiers
FT	CDS	10..927
FT		/*tag= a
FT	sig_peptide	10..72
FT		/*tag= b
FT	mat_peptide	73..924
FT		/*tag= c

XX
PN EP508358-A.
XX
PD 14-OCT-1992.
XX
PF 07-APR-1992; 92EP-0105956.
XX
PR 12-APR-1991; 91EP-0870062.
XX
PA (PROC) PROCTER & GAMBLE CO.
XX
PI Busch A, MacCorquodale F;
XX
DR WPI; 1992-341667/42.
DR P-PSDB; AAR27968.
XX
PT Laundry detergent compsns. - contg. alkali cellulase and PVP for
PT fabric colour maintenance upon laundering
XX
PS Disclosure; Page 14; 23pp; English.
XX

```
CC   The sequence is that of cDNA encoding an endoglucanase (alkaline
CC   cellulase) from Humicola insolens which is used with a polyvinyl-
CC   pyrrolidone (PVP) in a laundry detergent compsn. The action of the PVP
CC   and the endoglucanase in fabric colour maintenance upon laundering is
CC   superior to the sum of the individual actions of both these ingredients.
CC   See also AAQ29935.
XX
SQ   Sequence 1060 BP; 190 A; 377 C; 288 G; 205 T; 0 other;

  Query Match             37.3%;  Score 437.4;  DB 13;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 2.4e-70;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | ||'|   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||       ||||||  | ||||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | ||||||  | | | |||| ||||||  |    |||||    ||||||||  |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           ||||||||||||  || ||||| |||||||'|| |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
           || |||||| |    ||  |    ||||||||||||||||||| ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           ||||||| || || |||||||  ||||| || ||||| || ||||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| |||||||||||||| |||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
           | ||||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | |||||||||||||| | ||||| || |||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| ||||  ||||  |||| ||||| |||||||| ||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   |||||||||||  || |||||| | | ||     |||  | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                  |  || |   |   ||||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||          ||| ||||||||  || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| ||||  ||||  ||||| |||||||||| ||| |||||| ||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || |||| |||   |||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 10
AAQ30067
ID   AAQ30067 standard; DNA; 1060 BP.
XX
AC   AAQ30067;
XX
DT   02-APR-1993  (first entry)
XX
DE   Sequence encoding ~ 43 kD endogulcanase.
XX
KW   Detergent addative; protease; enzyme; endoglucanase; ss.
XX
OS   Humicola insolens DSM 1800.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     10..72
FT                   /*tag=  a
FT   mat_peptide     73..927
FT                   /*tag=  b
XX
PN   WO9218599-A.
XX
PD   29-OCT-1992.
XX
PF   10-APR-1992;   92WO-DK00116.
XX
PR   22-APR-1991;   91DK-0000737.
XX
PA   (NOVO ) NOVO-NORDISK AS.
XX
PI   Tsuchiya R,  Wagner P;
XX
DR   WPI; 1992-382092/46.
DR   P-PSDB; AAR28295.
XX
PT   Detergent additive contg. cellulase and specific protease - which
PT   does not degrade the cellulase during storage and clarifies the
PT   colour of dyed cellulosic materials
XX
PS   Disclosure; Page 8-9; 15pp; English.
XX
CC   The ~43 kD endoglucanase derived from Humicola insolens, DSM 1800 is
CC   described in detail in co-pending Danish patent application No. DK
CC   1159/90.  As a detergent additive, it has a higher degree of specificity
CC   then Bacillus lentus serine protease.  The term "higher degree of
CC   specificity" is defined as a protease which conditions degrades human
CC   insulin to fewer components.
XX
SQ   Sequence 1060 BP; 190 A; 377 C; 288 G; 205 T; 0 other;

  Query Match             37.3%;  Score 437.4;  DB 13;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 2.4e-70;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||       ||||||  | ||||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | ||||||  | | | |||| ||||||  |    |||||    ||||||||  |
```

```
Db    179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy    292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351

Db    239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy    352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411

Db    299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy    412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471

Db    359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy    472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531

Db    419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy    532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591

Db    479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy    592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651

Db    539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy    652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711

Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771

Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy    772 GCAACGG----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813

Db    719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy    814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870

Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930

Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961

Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 11
AAQ41732
ID   AAQ41732 standard; DNA; 1060 BP.
XX
AC   AAQ41732;
XX
DT   25-AUG-1993  (first entry)
XX
DE   Dye transfer inhibiting compsn. cellulase gene.
XX
KW   Detergent; homogeneous endoglucanase component; 43kD cellulase; ss.
XX
OS   Humicola insolens.
XX
FH   Key             Location/Qualifiers
FT   CDS             10..927
FT                   /*tag=  a
FT   sig_peptide     10..72
FT                   /*tag=  b
FT   mat_peptide     73..927
FT                   /*tag=  c
XX
. . . .

PN   EP540784-A.
XX
PD   12-MAY-1993.
XX
PF   06-NOV-1991;   91EP-0202882.
XX
PR   06-NOV-1991;   91EP-0202882.
XX
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Busch A,  McCorquodale F;
XX
DR   WPI; 1993-153868/19.
DR   P-PSDB; AAR37150.
XX
PT   Compsn. for inhibiting dye transfer during fabric washing - contains
PT   peroxidase, hydrogen peroxide, substrate and cellulase, esp. endo-
PT   glucanase from Humicola insolens
XX
PS   Disclosure; Page 18-19; 28pp; English.
XX
CC   The sequence is that encoding a cellulase which is characterised in
CC   that the cellulase provides at least 10% removal of immobilised
CC   radioactive labelled carboxymethyl cellulose according to the C14CMC
CC   method at 25 x 10(-6)% by weight of the cellulase protein in the
CC   test solution. It can be used as part of a compsn. for inhibiting
CC   dye transfer.
XX
SQ   Sequence 1060 BP; 190 A; 377 C; 288 G; 205 T; 0 other;

  Query Match             37.3%;  Score 437.4;  DB 14;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 2.4e-70;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy     55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114

Db      5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy    115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174

Db     65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy    175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234

Db    119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy    235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291

Db    179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy    292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351

Db    239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy    352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411

Db    299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy    412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471

Db    359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy    472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531

Db    419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy    532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591

Db    479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy    592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651

Db    539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598
```

```
Qy    652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
          ||||||||||  ||||  ||||  ||||  ||||| |||||||| ||||||  |||| || |
Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
          || |||| |||   ||||||||||||  || ||||||  | | ||     |||   | |
Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy    772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
          ||| | |                 |  || |   |  |||||| ||  |  |    |
Db    719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy    814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
          || ||| | |       ||         ||| |||||||  || ||| ||||||||||
Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
          |||| ||||  ||||  ||||| |||||||||| ||| |||||| ||| |||  ||||
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
          | || ||||  |||   |||||||| || ||
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929
```

```
RESULT 12
AAQ49941
ID   AAQ49941 standard; DNA; 1060 BP.
XX
AC   AAQ49941;
XX
DT   28-APR-1994  (first entry)
XX
DE   Endoglucanase enzyme.
XX
KW   Detergent; fabric; surfactant; softening clay; cellulase;
KW   Humicola insolens; Fusarium oxysporum; endoglucanase; ss.
XX
OS   Humicola insolens.
XX
FH   Key             Location/Qualifiers
FT   CDS             10..927
FT                   /*tag= a
FT   sig_peptide     10..72
FT                   /*tag= b
FT   mat_peptide     73..924
FT                   /*tag= c
XX
PN   AU9211048-A.
XX
PD   02-SEP-1993.
XX
PF   18-FEB-1992;   92AU-0011048.
XX
PR   18-FEB-1992;   92AU-0011048.
XX
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Baeck AC,  Busch A,  Convents AC;
XX
DR   WPI; 1993-328419/42.
DR   P-PSDB; AAR42063.
XX
PT   Detergent compsns., esp. for fabrics - contain surfactant,
PT   softening clay and high activity cellulase, partic. from Humicola
PT   insolens
XX
PS   Claim 10; Page 54-55; 71pp; English.
XX
CC   A new detergent comprises a high activity cellulase in combination
CC   with a softening clay.  The cellulase may be an endoglucanase
CC   enzyme derived from Humicola insolens (AAQ49941) or Fusarium oxysporum
CC   (AAQ49942).  The combination provides a higher than additive softening
CC   performance and excellent colour rejuvenation and whiteness
CC   maintenance for fabrics.
XX
SQ   Sequence 1060 BP; 190 A; 377 C; 288 G; 205 T; 0 other;
```

```
  Query Match             37.3%;  Score 437.4;  DB 14;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 2.4e-70;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy     55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
          ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db      5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy    115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
          ||   || ||   ||||||      ||||||  | ||||||||||||||||||| ||||
Db     65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy    175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
          |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db    119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy    235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
          ||||  |  | ||||||  | |  | |||| ||||||  |    |||||    ||||||||  |
Db    179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy    292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
          |||||||||||||  || ||||| |||||||||  |  ||||   |||  ||  ||  ||  ||
Db    239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy    352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
            || ||||||  |    ||  |    ||||||||||||||||||  ||||||||||  |
Db    299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy    412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
          |||||||  ||  ||  |||||||  |||||  ||  |||||  ||  ||||||||||  ||  ||  |
Db    359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy    472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
          |  ||  |||||  ||||||  ||  |||  ||||||||||||||  |||||||||  ||||  ||||
Db    419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy    532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
            |  ||||||||||  ||  ||||||    |  ||||||||||  |||||  |||  ||  |||
Db    479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy    592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
          ||||  ||||||  |  ||||||||||||||  | |||||  ||  |||||||||  |||
Db    539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy    652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
          |||||||||| ||||  ||||  ||||  ||||| |||||||| ||||||  |||| || |
Db    599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy    712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
          || |||| |||   ||||||||||||  || ||||||  | | ||     |||   | |
Db    659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy    772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
          ||| | |                 |  || |   |  |||||| ||  |  |    |
Db    719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy    814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
          || ||| | |       ||         ||| |||||||  || ||| ||||||||||
Db    779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy    871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
          |||| ||||  |||  ||||| |||||||||| ||| |||||| ||| |||  ||||
Db    839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy    931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
          | || ||||  |||   |||||||| || ||
Db    899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929
```

```
RESULT 13
AAT10182
ID   AAT10182 standard; DNA; 1060 BP.
XX
AC   AAT10182;
XX
DT   03-APR-1996  (first entry)
XX
DE   Alkaline endoglucanase Carezyme gene.
XX
KW   Alkaline endoglucanase; carezyme; cellulase; host cell;
KW   Fusarium graminearum; heterologous gene expression; ss.
XX
OS   Humicola insolens DSM 1800.
XX
FH   Key             Location/Qualifiers
FT   CDS             10..927
FT                   /*tag=  a
FT                   /product= pre-Carezyme
FT   sig_peptide     10..72
FT                   /*tag=  b
FT   mat_peptide     73..924
FT                   /*tag=  c
XX
PN   WO9600787-A1.
XX
PD   11-JAN-1996.
XX
PF   15-JUN-1995;   95WO-US07743.
XX
PR   15-MAR-1995;   95US-0404678.
PR   30-JUN-1994;   94US-0269449.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
XX
PI   Moyer DL,  Royer JC,  Shuster JR,  Yoder W;
XX
DR   WPI; 1996-077498/08.
DR   P-PSDB; AAR88471.
XX
PT   Non-toxic, non-toxigenic, non-pathogenic recombinant Fusarium host
PT   cell - used to produce heterologous proteins, pref. enzymes,
PT   hormones, growth factors or receptors
XX
PS   Disclosure; Page 22-24; 38pp; English.
XX
CC   The Humicola insolens DSM 1800 gene (AAT10182) coding for alkaline
CC   endoglucanase Carezyme (AAR88471), a derivative of a 43 kDa
CC   cellulase, was placed under the control of the promoter and
CC   terminator (see AAT10184-85) of the Fusarium oxysporum trypsin-like
CC   protease SP387 gene in vector pDM151.  Transformation of
CC   non-toxic, non-toxigenic, non-pathogenic, protease-deficient
CC   Fusarium graminearum ATCC 20334 with the plasmid allowed prodn.
CC   of Carezyme at levels of 6.0 g/l host cell; over 90% of
CC   secreted protein was Carezyme.
XX
SQ   Sequence 1060 BP; 190 A; 377 C; 288 G; 205 T; 0 other;

  Query Match             37.3%;  Score 437.4;  DB 17;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 2.4e-70;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           ||| ||| ||  |||||| |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | |||||| | | | |||| |||||| |    |||||    |||||||| |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||| || ||||| |||||||| |  ||||   ||| || || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||| |    ||  |    |||||||||||||||||| ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           ||||||| || || |||||||  ||||| || ||||| || |||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| |||||||||||||| ||||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
             | |||||||||| || ||||||    |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | |||||||||||||| | ||||| || ||||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| ||||  ||||  ||||  ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   |||||||||||  || |||||| | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG------------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                  |  || |   |  ||||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |        ||          ||| ||||||||  || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| ||||  ||||  ||||| |||||||||| ||| |||||| ||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || ||||  |||   |||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 14
AAV16102
ID   AAV16102 standard; cDNA; 1060 BP.
XX
AC   AAV16102;
XX
DT   21-JUL-1998  (first entry)
XX
DE   Humicola insolens surface-active endoglucanase gene.
XX
KW   endoglucanase; surface-active; cellulase; detergent; cleaning;
KW   performance; stain removal; soften; feel; colour; ss.
XX
OS   Humicola insolens.
XX
FH   Key             Location/Qualifiers
FT   CDS             10..927
```

```
FT                   /*tag=  a
FT                   /product= "endoglucanase"
FT   mat_peptide     73..924
FT                   /*tag=  b
FT   sig_peptide     10..72
FT                   /*tag=  c
XX
PN   WO9804663-A1.
XX
PD   05-FEB-1998.
XX
PF   25-JUL-1997;    97WO-US13194.
XX
PR   30-JUL-1996;    96US-0023125.
XX
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Boyer SL;
XX
DR   WPI; 1998-130664/12.
DR   P-PSDB; AAW46616.
XX
PT   Detergent composition containing both surface-active and non-surface
PT   active cellulase - softens and improves feel of cotton fabrics
PT   without causing loss in weight or tensile strength
XX
PS   Disclosure; Pages 41-43; 68pp; English.
XX
CC   The sequence is that encoding the endoglucanase enzyme of a
CC   surface-active cellulase produced by Humicola. The enzyme can be
CC   used in a detergent composition with a non-surface-active
CC   cellulase. This combination improves cleaning performance
CC   (maintains colours and removes stains), and softens and
CC   improves the feel of cotton fabrics without causing losses in
CC   weight or tensile strength.
XX
SQ   Sequence 1060 BP; 190 A; 377 C; 288 G; 205 T; 0 other;
```

Query Match 37.3%; Score 437.4; DB 19; Length 1060;
Best Local Similarity 70.4%; Pred. No. 2.4e-70;
Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

```
Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | ||||||  | | | |||| ||||||  |    |||||    |||||||| |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||||  || ||||| |||||||||  |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
           || |||||| |    ||  |    ||||||||||||||||||| ||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           |||||||  || || |||||||  ||||| || ||||| || |||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| |||||| ||  ||| |||||||||||||| ||||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
           |  |||||||||| || ||||||   |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | |||||||||||||| | ||||| || |||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| |||| |||| |||| ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||  |||||||||||  || |||||| | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                 |  || |   |  ||||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||         ||| ||||||||  || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| ||||  ||||  ||||| |||||||||| ||| |||||| ||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || ||||  |||   |||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929
```

```
RESULT 15
AAV15065
ID   AAV15065 standard; DNA; 1060 BP.
XX
AC   AAV15065;
XX
DT   19-MAY-1998  (first entry)
XX
DE   Humicola insolens EG V 43 KDa endoglucanase encoding DNA.
XX
KW   Humicola insolens; endoglucanase; isolation; microorganism;
KW   identification; ss.
XX
OS   Humicola insolens.
XX
FH   Key             Location/Qualifiers
FT   CDS             10..927
FT                   /*tag=  a
FT                   /product= "endoglucanase"
FT   sig_peptide     10..72
FT                   /*tag=  b
FT   mat_peptide     73..924
FT                   /*tag=  c
XX
PN   WO9743409-A2.
XX
PD   20-NOV-1997.
XX
PF   12-MAY-1997;    97WO-DK00216.
XX
PR   10-MAY-1996;    96DK-0000562.
XX
PA   (NOVO ) NOVO-NORDISK AS.
XX
PI   Dalboge H,  Diderichsen B,  Kauppinen S,  Sandal T;
XX
DR   WPI; 1998-008878/01.
DR   P-PSDB; AAW44266.
XX
```

PT Isolating novel DNA sequences from microorganisms - without the need
PT for culturing the microorganism
XX
PS Example 3; Page 47-48; 72pp; English.
XX
CC The present sequence encodes a protein with endoglucanase activity used
CC in an example of the present invention. The present invention describes
CC a novel method for providing a novel DNA sequence encoding a polypeptide
CC from a microorganism with an activity of interest. The method comprises:
CC (i) PCR amplification of the DNA with PCR primers with homology to (a)
CC known gene(s) encoding a polypeptide with an activity of interest; (ii)
CC linking the obtained PCR product of a 5' structural gene sequence and a
CC 3' structural gene sequence; (iii) expressing the resulting hybrid DNA
CC sequence; (iv) screening for hybrid DNA sequences encoding a polypeptide
CC with the activity of interest or a related activity; and (v) isolating
CC the hybrid DNA sequence identified in step (iv). This method provides
CC for identification and isolation of sequences from microorganisms
CC without having to cultivate and isolate the microorganism.
XX
SQ Sequence 1060 BP; 190 A; 377 C; 288 G; 205 T; 0 other;

Query Match 37.3%; Score 437.4; DB 19; Length 1060;
Best Local Similarity 70.4%; Pred. No. 2.4e-70;
Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

```
Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | |||||| | | | |||| |||||| |    |||||    ||||||||
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||||  || ||||| |||||||| |  ||||   ||| || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
           || ||||||  |    ||  |    ||||||||||||||||| ||||||||||||  |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           |||||||  ||  ||  |||||||  |||||  ||  |||||  ||  |||||||||||  ||  ||  |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           |  ||  |||||  ||||||  ||  |||  ||||||||||||||  |||||||||  ||||  ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
           |  ||||||||||  ||  ||||||   |  ||||||||||  |||||  |||  ||  |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   |  ||||||||||||||  | |||||  ||  |||||||||  |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           ||||||||||  ||||  ||||  ||||  |||||  ||||||||  ||||||  ||||  ||  |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           ||  ||||  |||   ||||||||||||  ||  ||||||  |  |  ||     |||   |  |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG------------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           |||  |  |                  |  ||  |   |  |||||||  ||  |  |   |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           ||  |||  |  |       ||          |||  ||||||||  ||  |||  ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           ||||  ||||  ||||  |||||  ||||||||||  |||  ||||||  |||  |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | ||  ||||  |||   ||||||||  ||  ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929
```

Search completed: February 27, 2003, 18:42:42
Job time : 311 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: February 27, 2003, 18:29:56 ; Search time 1642 Seconds
(without alignments)
11579.477 Million cell updates/sec

Title: US-10-007-521-11
Perfect score: 1174
Sequence: 1 GAGCAGCACCCCTCAAGCTG.....AAAAAAAAAAAAAAAAAAAA 1174

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	213	18.1	691	10	BE585661	BE585661 EST#6PSP6
	2	130.4	11.1	450	13	BI200729	BI200729 o1f05fs.r
	3	114.4	9.7	444	13	BI190695	BI190695 i3g10fs.r
	4	106.4	9.1	426	13	BI187295	BI187295 a1h11fs.r
	5	70.2	6.0	168	14	BQ152857	BQ152857 NF025H04I
c	6	69.6	5.9	203	9	AI620089	AI620089 ty48d08.x
	7	68.8	5.9	493	12	BG656882	BG656882 df25g08.y
	8	68.4	5.8	441	13	BM573394	BM573394 fx34f05.y
	9	68.4	5.8	596	14	BQ186959	BQ186959 UI-E-EJ1-
	10	68	5.8	179	9	AA237797	AA237797 mx14e11.r
c	11	68	5.8	369	9	AI870238	AI870238 wl20e09.x
c	12	68	5.8	442	9	AI721722	AI721722 fc31e12.x
c	13	67.6	5.8	271	10	AW193007	AW193007 x168h01.x
c	14	67.6	5.8	424	14	BQ391956	BQ391956 NISC_mq21
	15	67.6	5.8	441	12	BG897070	BG897070 HOA45-1-C
c	16	67.6	5.8	451	14	BQ525205	BQ525205 NISC_no09
c	17	67.6	5.8	476	13	BI496243	BI496243 df123h03.
c	18	67.6	5.8	542	14	BQ522269	BQ522269 NISC_nl15
c	19	67.6	5.8	543	14	BQ521817	BQ521817 NISC_nl13
c	20	67.6	5.8	546	14	BQ520637	BQ520637 NISC_nl06
c	21	67.6	5.8	550	14	BQ396262	BQ396262 NISC_ng19
	22	67.4	5.7	273	9	AU175573	AU175573 AU175573
c	23	67.4	5.7	435	9	AA806378	AA806378 oc22g02.s
	24	67.2	5.7	191	14	BQ942645	BQ942645 AGENCOURT
	25	67.2	5.7	201	9	AL712405	AL712405 DKFZp686N
	26	67.2	5.7	215	14	BQ932982	BQ932982 AGENCOURT
	27	67.2	5.7	246	9	AL598823	AL598823 DKFZp313K
	28	67.2	5.7	400	10	AV760389	AV760389 AV760389
	29	67.2	5.7	400	10	AV760391	AV760391 AV760391
	30	67.2	5.7	640	12	BF342223	BF342223 602012991
c	31	67	5.7	248	9	AA738097	AA738097 nx11h10.s
	32	67	5.7	295	9	AA268010	AA268010 va89d02.r
	33	67	5.7	315	9	AA270150	AA270150 va64h01.r
c	34	67	5.7	353	10	AW335461	AW335461 S47E2 AGS
	35	67	5.7	376	12	BG653511	BG653511 sad58c07.
	36	67	5.7	379	9	AA023416	AA023416 mh72a07.r
c	37	67	5.7	399	10	AW088521	AW088521 xd10c12.x
	38	67	5.7	435	14	BQ043166	BQ043166 UI-M-EM0-
c	39	67	5.7	479	13	BM114912	BM114912 L0812F05-
c	40	67	5.7	501	13	BM116326	BM116326 L0834D05-
	41	67	5.7	698	12	BF136434	BF136434 601783964
	42	67	5.7	740	12	BG294200	BG294200 602391273
	43	67	5.7	786	12	BG297249	BG297249 602394959
c	44	67	5.7	1279	12	BE784196	BE784196 601471421
	45	66.8	5.7	215	13	BI190568	BI190568 i2e11fs.r

ALIGNMENTS

RESULT 1
BE585661
LOCUS BE585661 691 bp mRNA linear EST 17-AUG-2000
DEFINITION EST#6PSP6_D02_d2_014 KSU wheat Fusarium graminearum infected spike cDNA library Triticum aestivum cDNA clone EST#6PSP6_D02_d2_014, mRNA sequence.
ACCESSION BE585661
VERSION BE585661.1 GI:9838604
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 691)
AUTHORS Fellers,J.P., Li,W.L., Hill-Ambroz,K., Matthews,A. and Gill,B.S.
TITLE The structure and function of the expressed portion of the wheat genomes - Kansas State Unversity. Fusarium graminearum infected spike cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: John Fellers
US Department of Agriculture, Agriculture Research Service, Plant Science and Entomology Unit
Dept. of Plant Pathology, 4006 Throckmorton Hall, Kansas State University, Manhattan, KS 66506, USA
Tel: 785-532-2367
Fax: 785-532-6167
Email: jpf@alfalfa.ksu.edu
Sequence have been trimmed to remove vector sequence and low

```
          quality sequence with phred score less than 20
          Seq primer: SP6.
FEATURES             Location/Qualifiers
     source          1. .691
                     /organism="Triticum aestivum"
                     /cultivar="Sumai3"
                     /db_xref="taxon:4565"
                     /clone="EST#6PSP6_D02_d2_014"
                     /clone_lib="KSU wheat Fusarium graminearum infected spike
                     cDNA library"
                     /tissue_type="Spike"
                     /dev_stage="Adult plant"
                     /lab_host="E. coli JM109"
                     /note="Vector: pGEM-T easy; Site_1: SacII; Site_2: SpeI;
                     Plants were grown in the greenhouse. Spikes were sprayed
                     with  Fusarium graminearum (at what stage). Total RNA, and
                     poly(A) RNA  were prepared from infected spikes. cDNA was
                     prepared using the  SmartTM PCR cDNA synthesis kit from
                     Clonetech. cDNA was cloned  into the pGEM-T easy vector
                     from Promega."
BASE COUNT      135 a    196 c    163 g    197 t
ORIGIN

  Query Match             18.1%;  Score 213;  DB 10;  Length 691;
  Best Local Similarity   64.9%;  Pred. No. 8.5e-20;
  Matches  364;  Conservative    0;  Mismatches  190;  Indels    7;  Gaps    3;

Qy      89 CCTGGCCGCTGCACTTCCTCTGGTCGCCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAG 148
           ||| |||   |     ||||| | ||     || ||   ||| ||||| || || ||  |
Db     131 CCTCGCCCTGGTCGGCCCTCTCGCCGTGAGTGCTGCTTCTGGAAGTGGTCACTCTACTCG 190

Qy     149 ATACTGGGACTGCTGCAAGCCGTCGTGCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACC 208
           ||||||||||||||||||||| || ||| |||||   || ||||||   ||||||   ||
Db     191 ATACTGGGACTGCTGCAAGCCTTCTTGCTCTTGGAGTGGCAAGGCCAAAGTCAGCGCCCC 250

Qy     209 GGTCTACGCGTGCGATGCCAACTTCCAGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTG 268
           |  |   | || ||    ||   | | | | |  || | || | |      || ||
Db     251 TGCTTTGACTTGTGACAAAAAAGACAACCCCATCACTAACCTGAACGCCGTCAACGGTTG 310

Qy     269 ---CAACGGCGGCTCGGCCTACTCCTGCGCCGACCAGACTCCCTGGGCGGTGAACGACAA 325
              |  || || || ||||   | ||  |  || |  | || ||||| || |||||| |
Db     311 TGAGAGTGGTGGTTCTGCCTTTGCTTGTACTAACTACTCCCCTTGGGCTGTCAACGACGA 370

Qy     326 TCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTG 385
           || || ||||| ||| | || || |   | || || ||  | ||  ||   ||||||||
Db     371 CCTTGCTTACGGTTTCACTGCTACCAAGCTTGCTGGTGGCACTGAGACCAGCTGGTGCTG 430

Qy     386 CGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCA 445
           || ||||| || |||| |||||  ||||||||||    || ||||  ||| | || ||
Db     431 TGCTTGCTATGCTCTCACTTTCACGACCGGTCCCGTAAAGGGAAAGAAGATGATTGTCCA 490

Qy     446 GTCAACGAGCACTGGCGGCGACCTGGGAAGTAACCAGTTCGATATCGCCATGCCCGGCGG 505
           ||| || | ||||||  || || || ||   |||||| || ||  |    ||||||||||
Db     491 GTCCACCAACACTGGTGGTGATCTCGGTGATAACCACTTTGACCTTATGATGCCCGGCGG 550

Qy     506 CGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGTTCGGC---GGCCTCCCCGGCGCTCA 562
           || || || || ||| | || |||| |||  ||||||||     ||||   || |||||
Db     551 TGGTGTCGGTATATTCGATGGATGCACCTCTGAGTTCGGCAAGCCCCTCGGTGGTGCTCA 610

Qy     563 ATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCTTCCCCGCGCCGCTCAAGCCCGG 622
           |||||||||| || |  ||   ||| || ||    ||||||| || |||||||   ||
Db     611 GTACGGCGGCATCTCTTTCCGAAGCCA-TGTGACAGTTTCCCCGAGCTGCTCAAGGATGG 669

Qy     623 CTGCCAGTGGCGGTTTGACTG 643
           ||||| ||||| ||||||||
Db     670 TTGCCACTGGCGATTTGACTG 690


RESULT 2
BI200729
LOCUS       BI200729                450 bp    mRNA    linear   EST 10-JUL-2001
DEFINITION  o1f05fs.r1 Fusarium sporotrichioides Tri 10 overexpressed cDNA
            library Fusarium sporotrichioides cDNA clone o1f05fs 5', mRNA
            sequence.
ACCESSION   BI200729
VERSION     BI200729.1  GI:14666701
KEYWORDS    EST.
SOURCE      Fusarium sporotrichioides.
  ORGANISM  Fusarium sporotrichioides
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreales; mitosporic Hypocreales; Fusarium.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Ren,Q., Tag,A., Peplow,A., Lai,H., Kupfer,C., Peterson,A., Beremand
            ,M. and Roe,B.
  TITLE     Analysis of a Fusarium sporotrichioides EST database
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
            Department of Chemistry and Biochemistry
            Advanced Center for Genome Technology, University of Oklahoma
            620 Parrington Oval, Norman, OK 73019, USA
            Tel: 405 325 4912
            Fax: 405 325 7762
            Email: broe@ou.edu
            Contact Dr. Marian Beremand regarding clone availability Included
            is the best homolog from a blastx search of Genbank nr 04-09-01
            633 5e-66 gi|1170140|sp|P45699 PUTATIVE ENDOGLUCANASE TYPE
            KPRECURSOR (EN
            Seq primer: T3
            High quality sequence stop: 440.
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Fusarium sporotrichioides"
                     /strain="Tri 10"
                     /db_xref="taxon:5514"
                     /clone="o1f05fs"
                     /clone_lib="Fusarium sporotrichioides Tri 10 overexpressed
                     cDNA library"
                     /note="Vector: pBlueScript SK-; Site_1: EcoRI; Site_2:
                     XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
                     ; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT       90 a    125 c    102 g    133 t
ORIGIN

  Query Match             11.1%;  Score 130.4;  DB 13;  Length 450;
  Best Local Similarity   62.4%;  Pred. No. 1.2e-08;
  Matches  222;  Conservative    0;  Mismatches  131;  Indels    3;  Gaps    1;

Qy      89 CCTGGCCGCTGCACTTCCTCTGGTCGCCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAG 148
           ||| |||         |||||| | ||     || ||   ||| |||||||| || ||  |
Db      95 CCTCGCCCTGTTCGGTCCTCTCGCCGTGAGTGCTGCTTCTGGAAGTGGCCACTCTACTCG 154

Qy     149 ATACTGGGACTGCTGCAAGCCGTCGTGCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACC 208
           ||||||||||||||||||||| || ||| |||||  ||| |||||    ||||||   ||
Db     155 ATACTGGGACTGCTGCAAGCCTTCTTGCTCTTGGAGCGGCAAGGCTAAAGTCAGCGCCCC 214

Qy     209 GGTCTACGCGTGCGATGCCAACTTCCAGCGCCTGTCCGACTTCAA---TGTCCAGTCGGG 265
           |      | || ||   |||     | |   |  | || | ||   |||| |     |
Db     215 TGCTCTGACTTGTGACAACAAAGATAACCCTATCACTAACCTGAACGCTGTCAACGGTTG 274

Qy     266 CTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACCAGACTCCCTGGGCGGTGAACGACAA 325
                 || || || || ||  | ||| || || |  | || ||||| || |||||| |
Db     275 TGAGGGTGGTGGTTCTGCTTATGCTTGCACCAACTACTCCCCGTGGGCTGTCAACGACGA 334

Qy     326 TCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTG 385
           || || ||||| ||||| || || |   | || || ||    ||  ||   ||||||||
Db     335 CCTTGCTTACGGTTTCGCTGCTACCAAGCTTGCTGGTGGTAGTGAGGCCAGCTGGTGCTG 394

Qy     386 CGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCGTCGCCGGCAAGACAATGGTGG 441
           || ||||| || ||||||||||  ||||||||||    |||||||  ||| | |
Db     395 TGCTTGCTATGCTCTCACCTTCACGACCGGTCCCGTGAAGGGCAAGAAGATGATTG 450


RESULT 3
BI190695
LOCUS       BI190695                444 bp    mRNA    linear   EST 10-JUL-2001
DEFINITION  i3g10fs.r1 Fusarium sporotrichioides Tri 10 overexpressed cDNA
```

```
            library Fusarium sporotrichioides cDNA clone i3g10fs 5', mRNA
            sequence.
ACCESSION   BI190695
VERSION     BI190695.1  GI:14664374
KEYWORDS    EST.
SOURCE      Fusarium sporotrichioides.
  ORGANISM  Fusarium sporotrichioides
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreales; mitosporic Hypocreales; Fusarium.
REFERENCE   1  (bases 1 to 444)
  AUTHORS   Ren,Q., Tag,A., Peplow,A., Lai,H., Kupfer,C., Peterson,A., Beremand
            ,M. and Roe,B.
  TITLE     Analysis of a Fusarium sporotrichioides EST database
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
            Department of Chemistry and Biochemistry
            Advanced Center for Genome Technology, University of Oklahoma
            620 Parrington Oval, Norman, OK 73019, USA
            Tel: 405 325 4912
            Fax: 405 325 7762
            Email: broe@ou.edu
            Contact Dr. Marian Beremand regarding clone availability Included
            is the best homolog from a blastx search of Genbank nr 04-09-01
            576 3e-59 gi|1170140|sp|P45699 PUTATIVE ENDOGLUCANASE TYPE
            KPRECURSOR (EN
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .444
                     /organism="Fusarium sporotrichioides"
                     /strain="Tri 10"
                     /db_xref="taxon:5514"
                     /clone="i3g10fs"
                     /clone_lib="Fusarium sporotrichioides Tri 10 overexpressed
                     cDNA library"
                     /note="Vector: pBlueScript SK-; Site_1: EcoRI; Site_2:
                     XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
                     ; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT       85 a    127 c     92 g    140 t
ORIGIN

  Query Match             9.7%;  Score 114.4;  DB 13;  Length 444;
  Best Local Similarity   61.7%;  Pred. No. 1.7e-06;
  Matches  200;  Conservative    0;  Mismatches  121;  Indels    3;  Gaps    1;

Qy      89 CCTGGCCGCTGCACTTCCTCTGGTCGCCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAG 148
           ||| |||         |||||| | ||     || ||   ||| |||||||| || ||  |
Db     121 CCTCGCCCTGTTCGGTCCTCTCGCCGTGAGTGCTGCTTCTGGAAGTGGCCACTCTACTCG 180

Qy     149 ATACTGGGACTGCTGCAAGCCGTCGTGCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACC 208
           ||||||||||||||||||| || ||| |||||  ||| |||||    ||||||   ||
Db     181 ATACTGGGACTGCTGCAAGCCTTCTTGCTCTTGGAGCGGCAAGGCTAAAGTCAGCGCCCC 240

Qy     209 GGTCTACGCGTGCGATGCCAACTTCCAGCGCCTGTCCGACTTCAA---TGTCCAGTCGGG 265
           |      | || ||   |||     | |   |  |  || | ||   |||| |     |
Db     241 TGCTCTGACTTGTGACAACAAAGATAACCCTATCACTAACCTGAACGCTGTCAACGGTTG 300

Qy     266 CTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACCAGACTCCCTGGGCGGTGAACGACAA 325
           || || || || ||  | ||| || || |  | || ||||| || ||||||| |
Db     301 TGAGGGTGGTGGTTCTGCTTATGCTTGCACCAACTACTCCCCGTGGGCTGTCAACGACGA 360

Qy     326 TCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTG 385
           || || ||||| ||||| || || |   | || || ||    ||  ||   ||||||||
Db     361 CCTTGCTTACGGTTTCGCTGCTACCAAGCTTGCTGGTGGTAGTGAGGCCAGCTGGTGCTG 420

Qy     386 CGCCTGCTACGCGCTCACCTTCAC 409
           || ||||| || ||||||||||
Db     421 TGCTTGCTATGCTCTCACCTTCAC 444


RESULT 4
BI187295
LOCUS       BI187295                426 bp    mRNA    linear   EST 10-JUL-2001
DEFINITION  a1h11fs.r1 Fusarium sporotrichioides Tri 10 overexpressed cDNA
            library Fusarium sporotrichioides cDNA clone a1h11fs 5', mRNA
            sequence.
ACCESSION   BI187295
VERSION     BI187295.1  GI:14660974
KEYWORDS    EST.
SOURCE      Fusarium sporotrichioides.
  ORGANISM  Fusarium sporotrichioides
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreales; mitosporic Hypocreales; Fusarium.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   Ren,Q., Tag,A., Peplow,A., Lai,H., Kupfer,C., Peterson,A., Beremand
            ,M. and Roe,B.
  TITLE     Analysis of a Fusarium sporotrichioides EST database
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: a1h11fs.f1
            Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
            Department of Chemistry and Biochemistry
            Advanced Center for Genome Technology, University of Oklahoma
            620 Parrington Oval, Norman, OK 73019, USA
            Tel: 405 325 4912
            Fax: 405 325 7762
            Email: broe@ou.edu
            Contact Dr. Marian Beremand regarding clone availability Included
            is the best homolog from a blastx search of Genbank nr 04-09-01
            565 5e-58 gi|1170140|sp|P45699 PUTATIVE ENDOGLUCANASE TYPE
            KPRECURSOR (EN
            Seq primer: T3
            High quality sequence stop: 338.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Fusarium sporotrichioides"
                     /strain="Tri 10"
                     /db_xref="taxon:5514"
                     /clone="a1h11fs"
                     /clone_lib="Fusarium sporotrichioides Tri 10 overexpressed
                     cDNA library"
                     /note="Vector: pBlueScript SK-; Site_1: EcoRI; Site_2:
                     XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
                     ; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT       82 a    119 c     91 g    134 t
ORIGIN

  Query Match             9.1%;  Score 106.4;  DB 13;  Length 426;
  Best Local Similarity   60.8%;  Pred. No. 2e-05;
  Matches  192;  Conservative    0;  Mismatches  121;  Indels    3;  Gaps    1;

Qy      89 CCTGGCCGCTGCACTTCCTCTGGTCGCCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAG 148
           ||| |||         |||||| | ||     || ||   ||| |||||||| || ||  |
Db     111 CCTCGCCCTGTTCGGTCCTCTCGCCGTGAGTGCTGCTTCTGGAAGTGGCCACTCTACTCG 170

Qy     149 ATACTGGGACTGCTGCAAGCCGTCGTGCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACC 208
           ||||||||||||||||||| || ||| |||||  ||| |||||    ||||||   ||
Db     171 ATACTGGGACTGCTGCAAGCCTTCTTGCTCTTGGAGCGGCAAGGCTAAAGTCAGCGCCCC 230

Qy     209 GGTCTACGCGTGCGATGCCAACTTCCAGCGCCTGTCCGACTTCAA---TGTCCAGTCGGG 265
           |      | || ||   |||     | |   |  |  || | ||   |||| |     |
Db     231 TGCTCTGACTTGTGACAACAAAGATAACCCTATCACTAACCTGAACGCTGTCAACGGTTG 290

Qy     266 CTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACCAGACTCCCTGGGCGGTGAACGACAA 325
           || || || || ||  | ||| || || |  | || ||||| || ||||||| |
Db     291 TGAGGGTGGTGGTTCTGCTTATGCTTGCACCAACTACTCCCCGTGGGCTGTCAACGACGA 350

Qy     326 TCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTG 385
           || || ||||| ||||| || || |   | || || ||    ||  ||   ||||||||
Db     351 CCTTGCTTACGGTTTCGCTGCTACCAAGCTTGCTGGTGGTAGTGAGGCCAGCTGGTGCTG 410

Qy     386 CGCCTGCTACGCGCTC 401
           || ||||| || |||
Db     411 TGCTTGCTATGCTCTC 426


RESULT 5
BQ152857
```

```
LOCUS       BQ152857                168 bp    mRNA    linear   EST 24-APR-2002
DEFINITION  NF025H04IR1F1044 Irradiated Medicago truncatula cDNA clone
            NF025H04IR 5', mRNA sequence.
ACCESSION   BQ152857
VERSION     BQ152857.1  GI:20289916
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 168)
  AUTHORS   Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula irradiated library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: May GD
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7391
            Fax: 580 221 7380
            Email: gdmay@noble.org
            Insert Length: 168   Std Error: 0.00
            Plate: 025  row: H  column: 04
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .168
                     /organism="Medicago truncatula"
                     /db_xref="taxon:3880"
                     /clone="NF025H04IR"
                     /clone_lib="Irradiated"
                     /tissue_type="seedlings"
                     /dev_stage="seedling"
                     /note="Vector: Lambda Zap; Seedlings were exposed either
                     to 100 gy gamma or 0.5, 1, 5, or  10 kJ/m2 UV irradiation.
                     Gamma-irradiated samples were harvested at 6,  12, 24 and
                     48 hours after treatment. UV-irradiated samples were
                     harvested  24 hours post-treatment. cDNA was prepared from
                     polyA+ enriched,  pooled samples of equivalent amounts of
                     total RNA from each sample.  The cDNA was directionally
                     ligated into the Uni-Zap XR vector  (Stratagene) and
                     packaged using the Gigapack III Gold packaging  extracts.
                     Phagemids containing cDNA inserts were in vivo excised
                     from  the recombinant Uni-ZAP XR vector using ExAssist
                     helper phage and the  E. coli strain XL1-Blue MRF'
                     (Stratagene). Excised plasmids were  plated using SOLR
                     cells."
BASE COUNT       85 a     16 c     24 g     43 t
ORIGIN

  Query Match             6.0%;  Score 70.2;  DB 14;  Length 168;
  Best Local Similarity   75.7%;  Pred. No. 2.3;
  Matches   87;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy   1060 TACATAGCACGCCGGTACACATCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTA 1119
           |||  |  | |  || | | ||    ||  || ||||| |  |   | ||  | |||| |
Db     54 TACTCATTATGTTGGAAAATATTATGCATGGATTTTGGTGATGTTTTTAGTTCTGTTTAA 113

Qy   1120 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    114 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 168


RESULT 6
AI620089/c
LOCUS       AI620089                203 bp    mRNA    linear   EST 07-MAR-2000
DEFINITION  ty48d08.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2282319 3',
            mRNA sequence.
ACCESSION   AI620089
VERSION     AI620089.1  GI:4629215
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 203)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 2710   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 163
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .203
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2282319"
                     /clone_lib="NCI_CGAP_Ut2"
                     /tissue_type="moderately-differentiated endometrial
                     adenocarcinoma, 3 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.85 kb. Life Technologies catalog #:
                     11539-012"
BASE COUNT       35 a     48 c     33 g     87 t
ORIGIN

  Query Match             5.9%;  Score 69.6;  DB 9;  Length 203;
  Best Local Similarity   81.0%;  Pred. No. 2.5;
  Matches   81;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy   1075 TACACATCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAA 1134
           || | | | |   ||  ||||||||||    |    ||  ||||||||||||||||||
Db    126 TAAAAACCCCGGGCCCCCTTTGGGGGCCCCTTTTCCCCTTTTTTAAAAAAAAAAAAAAA 67

Qy   1135 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
           ||||||||||||||||||||||||||||||||||||||||
Db     66 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 27


RESULT 7
BG656882
LOCUS       BG656882                493 bp    mRNA    linear   EST 26-APR-2001
DEFINITION  df25g08.y1 Wellcome CRC pRN3 St10 5 Xenopus laevis cDNA clone
            IMAGE:3558351 5', mRNA sequence.
ACCESSION   BG656882
VERSION     BG656882.1  GI:13796412
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
```

```
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B.
            Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington
            University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40RP from Gibco
            High quality sequence stop: 470.
FEATURES             Location/Qualifiers
     source          1. .493
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:3558351"
                     /clone_lib="Wellcome CRC pRN3 St10 5"
                     /tissue_type="embryo, stage 10.5"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
                     were oligo-dT primed and directionally cloned. Staging
                     according to  Nieuwkoop and Faber. Library was constructed
                     by N. Garrett, P. LeMaire, A.M. Zorn,  and J.B. Gurdon
                     (Wellcome/CRC Institute). "
BASE COUNT      190 a     82 c     67 g    133 t     21 others
ORIGIN

  Query Match             5.9%;  Score 68.8;  DB 12;  Length 493;
  Best Local Similarity   57.5%;  Pred. No. 2;
  Matches  103;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy     996 AGACCGTCAACTTCGTCATGCATATTTTTTGAGCGCTCAATACATACATAACCTTCGATT 1055
           | ||| | ||     |||    ||||||| |||  | |    ||| |||   || || |
Db     310 AAACCTTAAATCAAATCACAATTATTTTTCTAGCCTTTAGATTATATATATTTTTTGACT 369

Qy    1056 CTTGTACATAGCACGCCGGTACACATCTCACACCGACTTTGGGGGCGGAATCAGGCCCGT 1115
           |    |    ||   |  ||  ||||||      ||
Db     370 GTCTCGCCATACATCACATTAATCATCTCTGGTTCACAGGNNANAAANNNNNNNNNNNNN 429

Qy    1116 TTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     430 NNANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 488


RESULT 8
BM573394
LOCUS       BM573394                441 bp    mRNA    linear   EST 22-FEB-2002
DEFINITION  fx34f05.y1 Zebrafish SJD 15 day embryo Danio rerio cDNA clone
            5625345 5' similar to SW:RUXG_HUMAN Q15357 SMALL NUCLEAR
            RIBONUCLEOPROTEIN G ;, mRNA sequence.
ACCESSION   BM573394
VERSION     BM573394.1  GI:18861861
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library construction by: Joe Barnes and Steve Johnson. DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            distribution: RessourcenZentrumPrimarDatenbank, Berlin, Germany
            (web address: www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 432.
FEATURES             Location/Qualifiers
     source          1. .441
                     /organism="Danio rerio"
                     /strain="SJD"
                     /db_xref="taxon:7955"
                     /clone="5625345"
                     /clone_lib="Zebrafish SJD 15 day embryo"
                     /tissue_type="whole embryo, 15 day"
                     /lab_host="DH10B"
                     /note="Vector: pAMP1; Site_1: EcoRI; Site_2: NotI; First
                     strand cDNA synthesis was primed using oligo-dT on
                     magnetic beads  with an additional primer
                     5'-gcggccgctaatacgactcacta-taggg-3'. Second strand
                     synthesis  was a 3-cycle PCR using the primers
                     5'-ggccgctaatacgactcactatag-3' and 5'
                     -aagcagtggtaacaacgcagagtacttt-ttttttttttttvn-3'. cDNA was
                     subsequently amplified  in a 7-cycle PCR with the
                     following primers: 5'-ggccgctaatacgactcactatag-3'  and
                     5'-aagcagtggt-aacaacgcag. Deoxy-UMP adaptors were added in
                     a third PCR (5 cycles)  and the primers
                     5'-caucaucaucauggccgctaatacgactcactataggg-3' and
                     5'-cuacuacuacuaaagcagtggtaacaacgcagagtac-3'. Ends were
                     treated with uracil DNA glycosylase  and product with 3'
                     overhangs was annealed to complementary ends of pAMP1.
                     Insert can be excised  using EcoRI and NotI. Library
                     constructed by Joe Barnes and Steve  Johnson (Washington
                     University)."
BASE COUNT      200 a     60 c     64 g    117 t
ORIGIN

  Query Match             5.8%;  Score 68.4;  DB 13;  Length 441;
  Best Local Similarity   71.4%;  Pred. No. 2.4;
  Matches   90;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy    1049 TTCGATTCTTGTACATAGCACGCCGGTACACATCTCACACCGACTTTGGGGGCGGAATCA 1108
           ||  ||  ||   | | | |     |||||||  ||| |     ||| | |    ||| |
Db     279 TTATATATTTCCTCTTTGTATTTGTGTACACAATTCATAAAACATTTTGTGAATAAATGA 338

Qy    1109 GGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1168
           |   | || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     339 AGTGTGATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 398

Qy    1169 AAAAAA 1174
           ||||||
Db     399 AAAAAA 404


RESULT 9
BQ186959
LOCUS       BQ186959                596 bp    mRNA    linear   EST 30-APR-2002
DEFINITION  UI-E-EJ1-ajy-j-05-0-UI.r1 UI-E-EJ1 Homo sapiens cDNA clone
            UI-E-EJ1-ajy-j-05-0-UI 5', mRNA sequence.
ACCESSION   BQ186959
VERSION     BQ186959.1  GI:20362510
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 596)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
```

```
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 299-359, >POLY_A#Simple_repeat
            Seq primer: M13 REVERSE.
FEATURES              Location/Qualifiers
     source           1. .596
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="UI-E-EJ1-ajy-j-05-0-UI"
                      /clone_lib="UI-E-EJ1"
                      /tissue_type="fetal eyes, lens, eye anterior segment,
                      optic nerve, retina, Retina Foveal and Macular, RPE and
                      Choroid"
                      /dev_stage="fetal and adult"
                      /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                      /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                      modified polylinker; Site_1: EcoR I; Site_2: Not I;
                      UI-E-EJ1 is a subtracted cDNA library constructed
                      according to Bonaldo, Lennon and Soares, Genome Research,
                      6:791-806, 1996. First strand cDNA synthesis was primed
                      with an oligo-dT primer containing a Not I site. Double
                      stranded cDNA was ligated to an EcoR I adaptor, digested
                      with Not I, and cloned directionally into pT7T3-Pac
                      vector. The oligonucleotide used to prime the synthesis of
                      first-strand cDNA contains a library tag sequence that is
                      located between the Not I site and the (dT)18 tail. The
                      sequence tags for this library are: fetal eyes, AGAATCAAGA
                      ; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT;
                      optic nerve, CCATTAAGTG; retina, CCGCG; Retina Foveal and
                      Macular, GTCC; RPE and Choroid, ACCTA. This library was
                      created for the program, Gene Discovery in the Visual
                      System, supported by National Eye Institute (NEI)."
BASE COUNT      191 a    144 c    125 g    136 t
ORIGIN

  Query Match             5.8%;  Score 68.4;  DB 14;  Length 596;
  Best Local Similarity   69.4%;  Pred. No. 2;
  Matches   93;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy   1041 ACATAACCTTCGATTCTTGTACATAGCACGCCGGTACACATCTCACACCGACTTTGGGGG 1100
           | ||   ||||  ||||  |       |  ||    | ||||| || ||  ||  |
Db    220 ATATCTGCTTCCCTTCTGCTGTCCTCCCAGCTCAGAGACATCACAGACTCACACTTTCAT 279

Qy   1101 CGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1160
             | ||||    ||||| ||||||||||||||||||||||||||||||||||||||||
Db    280 TACAGTCAGCTATGTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 339

Qy   1161 AAAAAAAAAAAAAA 1174
           ||||||||||||||
Db    340 AAAAAAAAAAAAAA 353


RESULT 10
AA237797
LOCUS       AA237797                179 bp    mRNA    linear   EST 03-MAR-1997
DEFINITION  mx14e11.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:680204
            5', mRNA sequence.
ACCESSION   AA237797
VERSION     AA237797.1  GI:1861820
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 179)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:419908
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 166.
FEATURES              Location/Qualifiers
     source           1. .179
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:680204"
                      /clone_lib="Soares mouse NML"
                      /tissue_type="Liver"
                      /lab_host="DH10B"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTTTT 3'];
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      constructed and normalized by Bento Soares and M.Fatima
                      Bonaldo."
BASE COUNT       85 a     29 c     21 g     44 t
ORIGIN

  Query Match             5.8%;  Score 68;  DB 9;  Length 179;
  Best Local Similarity   80.0%;  Pred. No. 4.5;
  Matches   80;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy   1075 TACACATCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAA 1134
           | |||  |||  |  |||  ||   ||  || | | || | | |  ||||||||||||||||
Db     72 TCCACGCCTCTTATGGACATTAAAGGTCGATTGATGCACTTCTGAAAAAAAAAAAAAAAA 131

Qy   1135 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
           ||||||||||||||||||||||||||||||||||||||||
Db    132 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 171


RESULT 11
AI870238/c
LOCUS       AI870238                369 bp    mRNA    linear   EST 07-MAR-2000
DEFINITION  wl20e09.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2425480 3',
            mRNA sequence.
ACCESSION   AI870238
VERSION     AI870238.1  GI:5544206
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
```

```
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1141   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 363.
FEATURES             Location/Qualifiers
     source          1. .369
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2425480"
                     /clone_lib="NCI_CGAP_Ut1"
                     /tissue_type="well-differentiated endometrial
                     adenocarcinoma, 7 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Life Technologies catalog #:
                     11538-014"
BASE COUNT      105 a     97 c     81 g     86 t
ORIGIN

  Query Match             5.8%;  Score 68;  DB 9;  Length 369;
  Best Local Similarity   60.1%;  Pred. No. 3;
  Matches  113;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy     987 GGAGCAAGGAGACCGTCAACTTCGTCATGCATATTTTTTGAGCGCTCAATACATACATAA 1046
           || | || | | ||  ||||  |  ||  |    |||| |||   |  |  |    |
Db     188 GGGGGAAAGGGCCCCCCAACACCCCCAATCCAGGTTTTGGAGGCATTGAGCCCCCGACTT 129

Qy    1047 CCTTCGATTCTTGTACATAGCACGCCGGTACACATCTCACACCGACTTTGGGGGCGGAAT 1106
            ||| | || || |   | |  | ||  |   |  | | | ||  |  |  |||   ||
Db     128 TCTTGGGTTTTTTTTTTTTGTCCCCCCCTGGGCCCCCCCCCCCCGCCCTTTGGGAATAAA 69

Qy    1107 CAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1166
             | ||   |||     ||||||||||||||||||||||||||||||||||||||||||
Db      68 GGGCCCTCCTTTTGCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 9

Qy    1167 AAAAAAAA 1174
           ||||||||
Db       8 AAAAAAAA 1


RESULT 12
AI721722/c
LOCUS       AI721722                442 bp    mRNA    linear   EST 07-JUN-2001
DEFINITION  fc31el2.x1 Zebrafish WashU MPIMG EST Danio rerio cDNA clone
            IMAGE:3723022 3', mRNA sequence.
ACCESSION   AI721722
VERSION     AI721722.1  GI:5040051
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 442)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: fc31el2.y1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T7 ET from Amersham
            High quality sequence stop: 429.
FEATURES             Location/Qualifiers
     source          1. .442
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="IMAGE:3723022"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT3'];
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries.  Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT      125 a    106 c     95 g    116 t
ORIGIN

  Query Match             5.8%;  Score 68;  DB 9;  Length 442;
  Best Local Similarity   83.7%;  Pred. No. 2.7;
  Matches   77;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy    1083 TCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAAA 1142
           || | ||   |||  |    |  | ||||||||||| |||||||||||||||||||||||
Db      96 TCCCCCCTGTTTTTAGTAAAGGTTTAGGCCCGTTTAAAAAAAAAAAAAAAAAAAAAAA 37

Qy    1143 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
           ||||||||||||||||||||||||||||||
Db      36 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5


RESULT 13
AW193007/c
LOCUS       AW193007                271 bp    mRNA    linear   EST 29-NOV-1999
DEFINITION  xl68h01.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2679889 3'
            similar to contains element MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AW193007
VERSION     AW193007.1  GI:6471706
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 271)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Life Technologies catalog #: 11548-013
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 267.
FEATURES             Location/Qualifiers
     source          1. .271
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2679889"
                     /clone_lib="NCI_CGAP_Pan1"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT       51 a     71 c     56 g     93 t
ORIGIN

  Query Match             5.8%;  Score 67.6;  DB 10;  Length 271;
  Best Local Similarity   80.6%;  Pred. No. 4;
  Matches   79;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy   1077 CACATCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAA 1136
         | |  | | | ||  ||||||||||| |  |  | |||     ||||||||||||||||||
Db    130 CCCCCCCCCCCCCCCCTTTGGGGGCTGGTTGGGCCCCTGGGGAAAAAAAAAAAAAAAAAAA 71

Qy   1137 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
         |||||||||||||||||||||||||||||||||||||
Db     70 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 33


RESULT 14
BQ391956/c
LOCUS       BQ391956                424 bp    mRNA    linear   EST 22-MAY-2002
DEFINITION  NISC_mq21f07.x2 NICHD XGC Emb5 Silurana tropicalis cDNA clone
            IMAGE:5309316 3', mRNA sequence.
ACCESSION   BQ391956
VERSION     BQ391956.1  GI:21079643
KEYWORDS    EST.
SOURCE      western clawed frog.
  ORGANISM  Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   NIH-XCG http://image.llnl.gov/image/html/xenopuslib_info.shtml.
  TITLE     National Institute of Child Health and Human Development, National
            Cancer Insitute, Xenopus Gene Collection
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation:
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
             DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC)
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            info@image.llnl.gov
            Plate: LLAM11783  row: K  column: 13
            Seq primer: -21M13 forward primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .424
                     /organism="Silurana tropicalis"
                     /db_xref="taxon:8364"
                     /clone="IMAGE:5309316"
                     /clone_lib="NICHD XGC Emb5"
                     /tissue_type="gastrula"
                     /dev_stage="embryo, stages 10-13"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Vector: pCMV-SPORT6.ccdb; Site_1: NotI; Site_2:
                     EcoRV; Cloned unidirectionally. Primer: Oligo dT. Average
                     insert size 2.0 kb. Constructed by Invitrogen. Note: This
                     is a Xenopus Gene Collection (XGC) library."
BASE COUNT      117 a     76 c     68 g    163 t
ORIGIN

  Query Match             5.8%;  Score 67.6;  DB 14;  Length 424;
  Best Local Similarity   64.9%;  Pred. No. 3.1;
  Matches  100;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy   1019 ATTTTTTGAGCGCTCAATACATACATAACCTTCGATTCTTGTACATAGCACGCCGGTACA 1078
         | ||||| |   |       | | | || |||   | |||| |||||| |      |
Db    155 AGTTTTTTAATTCCGGGGGTAAAAAAAAACTTAAGTCCTTGAACATAGAAATGTAATTTT 96

Qy   1079 CATCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAA 1138
          | ||  ||   |   |||||| ||||   |         ||||||||||||||||||
Db     95 TAACTATCAATAAAACTGGGGGAGGAAGGGGAAAAAAAAAAAAAAAAAAAAAAAAAAAA 36

Qy   1139 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1172
          ||||||||||||||||||||||||||||||||||
Db     35 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2


RESULT 15
BG897070
LOCUS       BG897070                441 bp    mRNA    linear   EST 06-NOV-2001
DEFINITION  HOA45-1-C3.R HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA
            , mRNA sequence.
ACCESSION   BG897070
VERSION     BG897070.1  GI:14307311
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
            Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
            Lark,M.W.
  TITLE     Identification and initial characterization of 5000 expressed
            sequenced tags (ESTs) each from adult human normal and
            osteoarthritic cartilage cDNA libraries
  JOURNAL   Osteoarthr. Cartil. 9 (7), 641-653 (2001)
  MEDLINE   21482651
COMMENT     Contact: Sanjay Kumar
            UW2109
            GlaxoSmithKline
            709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
            Tel: 610-270-7245
            Fax: 610-270-5598
            Email: sanjay_kumar-1@gsk.com
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .441
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HOA (Human Osteoarthritic Cartilage)"
                     /tissue_type="cartilage"
                     /lab_host="E.coli DH10 B"
                     /note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
                     Directional"
BASE COUNT      177 a     94 c    106 g     64 t
ORIGIN
```

Query Match 5.8%; Score 67.6; DB 12; Length 441;
Best Local Similarity 77.4%; Pred. No. 3;
Matches 82; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Qy   1069 CGCCGGTACACATCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAA 1128
          ||||  | ||   | ||  ||| | | || | | ||   || | |  | | |||||||||
Db    268 CGCCACTGCAGTCCGCAGTCCGGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAAAA 327

Qy   1129 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    328 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 373
```

Search completed: February 27, 2003, 20:03:37
Job time : 1650 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: February 27, 2003, 17:41:10 ; Search time 137 Seconds
(without alignments)
5344.003 Million cell updates/sec

Title: US-10-007-521-11
Perfect score: 1174
Sequence: 1 GAGCAGCACCCCTCAAGCTG.....AAAAAAAAAAAAAAAAAAAA 1174

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 460893 seqs, 311809382 residues

Total number of hits satisfying chosen parameters: 921786

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	437.4	37.3	1060	10	US-09-735-787-1	Sequence 1, Appli
	2	353.6	30.1	936	7	US-08-841-636A-30	Sequence 30, Appl
	3	315.2	26.8	1473	10	US-09-735-787-3	Sequence 3, Appli
	4	83.2	7.1	887	7	US-08-841-636A-36	Sequence 36, Appl
	5	66	5.6	682	10	US-09-764-847-20	Sequence 20, Appl
	6	65.6	5.6	371	10	US-09-925-299-16	Sequence 16, Appl
	7	64.6	5.5	1341	10	US-09-764-846-41	Sequence 41, Appl
c	8	64	5.5	95	10	US-09-919-580-246	Sequence 246, App
	9	63.8	5.4	1908	10	US-09-745-763-69	Sequence 69, Appl
	10	63.6	5.4	917	10	US-09-925-301-29	Sequence 29, Appl
	11	63.4	5.4	1377	10	US-09-822-830A-588	Sequence 588, App
c	12	63.2	5.4	355	10	US-09-960-352-14757	Sequence 14757, A
c	13	63	5.4	1545	10	US-09-764-877-3172	Sequence 3172, Ap
c	14	62.6	5.3	391	10	US-09-960-352-14316	Sequence 14316, A
	15	62.6	5.3	1817	10	US-09-764-864-197	Sequence 197, App
	16	62.6	5.3	2203	9	US-10-224-562-1	Sequence 1, Appli
	17	62.6	5.3	2203	10	US-09-801-861-1	Sequence 1, Appli
	18	62.2	5.3	2379	9	US-10-036-041-71	Sequence 71, Appl
	19	62.2	5.3	2379	9	US-10-035-855-71	Sequence 71, Appl
	20	62.2	5.3	2379	9	US-10-174-590-483	Sequence 483, App
	21	62.2	5.3	2379	9	US-10-176-758-483	Sequence 483, App
	22	62.2	5.3	2379	9	US-10-175-737-483	Sequence 483, App
	23	62.2	5.3	2379	9	US-10-173-706-483	Sequence 483, App
	24	62.2	5.3	2379	9	US-10-175-738-483	Sequence 483, App
	25	62.2	5.3	2379	9	US-10-175-752-483	Sequence 483, App
	26	62.2	5.3	2379	9	US-10-176-482-483	Sequence 483, App
	27	62.2	5.3	2379	9	US-10-176-757-483	Sequence 483, App
	28	62.2	5.3	2379	9	US-10-176-913-483	Sequence 483, App
	29	62.2	5.3	2379	9	US-10-180-552-483	Sequence 483, App
	30	62.2	5.3	2379	9	US-10-180-557-483	Sequence 483, App
	31	62.2	5.3	2379	9	US-09-931-836-71	Sequence 71, Appl
	32	62.2	5.3	2379	9	US-10-173-700-483	Sequence 483, App
	33	62.2	5.3	2379	9	US-10-174-572-483	Sequence 483, App
	34	62.2	5.3	2379	9	US-10-174-579-483	Sequence 483, App
	35	62.2	5.3	2379	9	US-10-174-582-483	Sequence 483, App
	36	62.2	5.3	2379	9	US-10-174-588-483	Sequence 483, App
	37	62.2	5.3	2379	9	US-10-175-739-483	Sequence 483, App
	38	62.2	5.3	2379	9	US-10-175-740-483	Sequence 483, App
	39	62.2	5.3	2379	9	US-10-175-743-483	Sequence 483, App
	40	62.2	5.3	2379	9	US-10-176-488-483	Sequence 483, App
	41	62.2	5.3	2379	9	US-10-176-492-483	Sequence 483, App
	42	62.2	5.3	2379	9	US-10-176-747-483	Sequence 483, App
	43	62.2	5.3	2379	9	US-10-176-750-483	Sequence 483, App
	44	62.2	5.3	2379	9	US-10-176-985-483	Sequence 483, App
	45	62.2	5.3	2379	9	US-10-176-987-483	Sequence 483, App

ALIGNMENTS

```
RESULT 1
US-09-735-787-1
; Sequence 1, Application US/09735787
; Patent No. US20010036910A1
;  GENERAL INFORMATION:
;       APPLICANT: Rasmussen, Grethe
;                  Mikkelsen, Jan Moller
;                  Schulein, Martin
;                  Patkar, Shankant A.
;                  Hagen, Fred
;       TITLE OF INVENTION: A Cellulase Preparation Comprising an
;                           Endoglucanase Enzyme
;       NUMBER OF SEQUENCES: 33
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: No. US20010036910A1o No. US20010036910A1disk of No. US200100
;            STREET: 405 Lexington Avenue, 64th Floor
;            CITY: New York
;            STATE: New York
;            COUNTRY: United States of America
;            ZIP: 10174-6401
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/735,787
;            FILING DATE: 13-Dec-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/189,028
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Lambiris, Elias J.
;            REGISTRATION NUMBER: 33,728
;            REFERENCE/DOCKET NUMBER: 3469.214-US
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 212-867-0123
;            TELEFAX: 212-878-9655
;  INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: Humicola insolens
```

```
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;          MOLECULE TYPE: cDNA
;          HYPOTHETICAL: NO
;          ORIGINAL SOURCE:
;             STRAIN: DSM 1800
;          FEATURE:
;             NAME/KEY:  mat_peptide
;             LOCATION:  73..924
;          FEATURE:
;             NAME/KEY:  sig_peptide
;             LOCATION:  10..72
;          FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  10..924
;          SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-735-787-1

  Query Match             37.3%;  Score 437.4;  DB 10;  Length 1060;
  Best Local Similarity   70.4%;  Pred. No. 1.8e-72;
  Matches  655;  Conservative    0;  Mismatches  246;  Indels   30;  Gaps    4;

Qy      55 CCAGGATGCGCTCTACTCCCGTTCTTCGCACAACCCTGGCCGCTGCACTTCCTCTGGTCG 114
           ||| |||||| ||  | ||| | || |   |  || | |  || || || ||  || | |
Db       5 CCAAGATGCGTTCCTCCCCCCTCCTCCCGTCCGCCGTTGTGGCCGCCCTGCCGGTGTTGG 64

Qy     115 CCTCCGCGGCCAGTGGCAGTGGCCAGTCCACGAGATACTGGGACTGCTGCAAGCCGTCGT 174
           ||   || ||   ||||||      ||||||  | ||||||||||||||||||| ||||
Db      65 CCCTTGCCGCTGATGGCAG------GTCCACCCGCTACTGGGACTGCTGCAAGCCTTCGT 118

Qy     175 GCGCTTGGCCCGGGAAGGCCGCCGTCAGCCAACCGGTCTACGCGTGCGATGCCAACTTCC 234
           |||  ||| ||  ||||||  |||| | ||| || ||||   | ||| | ||||||||||
Db     119 GCGGCTGGGCCAAGAAGGCTCCCGTGAACCAGCCTGTCTTTTCCTGCAACGCCAACTTCC 178

Qy     235 AGCGCCTGTCCGACTTCAATGTCCAGTCGGGCTGCAA---CGGCGGCTCGGCCTACTCCT 291
           ||||  |  | ||||||| | | | |||| |||||| |    |||||    |||||||| |
Db     179 AGCGTATCACGGACTTCGACGCCAAGTCCGGCTGCGAGCCGGGCGGTGTCGCCTACTCGT 238

Qy     292 GCGCCGACCAGACTCCCTGGGCGGTGAACGACAATCTCGCCTACGGCTTCGCCGCGACGA 351
           |||||||||||| || ||||| |||||||||| |  ||||   ||| || || || || ||
Db     239 GCGCCGACCAGACCCCATGGGCTGTGAACGACGACTTCGCGCTCGGTTTTGCTGCCACCT 298

Qy     352 GCATCGCCGGCGGGTCCGAATCCTCGTGGTGCTGCGCCTGCTACGCGCTCACCTTCACTT 411
             || |||||| |    ||  |    |||||||||||||||||||| |||||||||||| |
Db     299 CTATTGCCGGCAGCAATGAGGCGGGCTGGTGCTGCGCCTGCTACGAGCTCACCTTCACAT 358

Qy     412 CCGGTCCCGTCGCCGGCAAGACAATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGG 471
           |||||| || || |||||||  ||||| || ||||| || |||||||||| || || |
Db     359 CCGGTCCTGTTGCTGGCAAGAAGATGGTCGTCCAGTCCACCAGCACTGGCGGTGATCTTG 418

Qy     472 GAAGTAACCAGTTCGATATCGCCATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCA 531
           | || ||||| ||||||  ||  ||| |||||||||||||| |||||||| |||| ||||
Db     419 GCAGCAACCACTTCGATCTCAACATCCCCGGCGGCGGCGTCGGCATCTTCGACGGATGCA 478

Qy     532 GCTCGCAGTTCGGCGGCCTCCCCGGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGT 591
             | |||||||||||  || ||||||    |  |||||||||| ||||| ||| || |||
Db     479 CTCCCCAGTTCGGCGGTCTGCCCGGCCAGCGCTACGGCGGCATCTCGTCCCGCAACGAGT 538

Qy     592 GCGATTCCTTCCCCGCGCCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGA 651
           |||||   |||||||   | ||||||||||||||| | ||||| || |||||||| |||
Db     539 GCGATCGGTTCCCCGACGCCCTCAAGCCCGGCTGCTACTGGCGCTTCGACTGGTTCAAGA 598

Qy     652 ACGCCGACAACCCGACGTTCACGTTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCC 711
           |||||||||| |||| ||||  |||| ||||| |||||||| |||||| |||| || |
Db     599 ACGCCGACAATCCGAGCTTCAGCTTCCGTCAGGTCCAGTGCCCAGCCGAGCTCGTCGCTC 658

Qy     712 GCTCCGGCTGCAAGCGCAACGACGACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCGGTG 771
           || |||| |||   ||||||||||||  || |||||| | | ||     |||   | |
Db     659 GCACCGGATGCCGCCGCAACGACGACGGCAACTTCCCTGCCGTCCAGATCCCCTCCAGCA 718

Qy     772 GCAACGG-----------------TGGCACCGGGACGCCCACGTCGACTGCGCCTGGGT 813
           ||| | |                 |  || |   |   ||||||| ||  |  |    |
Db     719 GCACCAGCTCTCCGGTCAACCAGCCTACCAGCACCAGCACCACGTCCACCTCCACCACCT 778

Qy     814 CGGGCCAGAC---GTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGT 870
           || ||| | |       ||        ||| ||||||||  || ||| ||||||||||
Db     779 CGAGCCCGCCAGTCCAGCCTACGACTCCCAGCGGCTGCACTGCTGAGAGGTGGGCTCAGT 838

Qy     871 GCGGTGGCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGT 930
           |||| ||||  ||||  ||||| |||||||||||| ||| |||||| ||| |||  ||||
Db     839 GCGGCGGCAATGGCTGGAGCGGCTGCACCACCTGCGTCGCTGGCAGCACTTGCACGAAGA 898

Qy     931 TGAACGACTACTACTCGCAGTGCCTCTAAAC 961
           | || ||||  |||   |||||||| || ||
Db     899 TTAATGACTGGTACCATCAGTGCCTGTAGAC 929


RESULT 2
US-08-841-636A-30
; Sequence 30, Application US/08841636A
; Patent No. US20020168751A1
; GENERAL INFORMATION:
;   APPLICANT:  Miettinen-Oinonen, Arja
;   APPLICANT:  Londesborough, John
;   APPLICANT:  Vehmaanper , Jari
;   APPLICANT:  Haakana, Heli
;   APPLICANT:  M ntyl , Arja
;   APPLICANT:  Lantto, Raija
;   APPLICANT:  Elovainio, Minna
;   APPLICANT:  Joutsjoki, Vesa
;   APPLICANT:  Paloheimo, Marja
;   APPLICANT:  Suominen, Pirkko
;   TITLE OF INVENTION:  NOVEL CELLULASES, THE GENES ENCODING THEM AND
;   TITLE OF INVENTION:  USES THEREOF
;   NUMBER OF SEQUENCES:  45
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;     STREET:  1100 New York Avenue, N.W., Suite 600
;     CITY:  Washington
;     STATE:  D.C.
;     COUNTRY:  USA
;     ZIP:  20005
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette, 3.50 inch
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/841,636A
;     FILING DATE:  30-APR-1997
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 60/005,335
;     FILING DATE:  17-OCT-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 60/007,926
;     FILING DATE:  04-DEC-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 60/020,840
;     FILING DATE:  28-JUN-1996
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/732,181
;     FILING DATE:  16-OCT-1996
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  PCT/FI96/00550
;     FILING DATE:  17-OCT-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Timothy J. Shea, Jr.
;     REGISTRATION NUMBER:  41,306
;     REFERENCE/DOCKET NUMBER:  1716.0510005/MAC/TJS
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (202)371-2600
;     TELEFAX:  (202)371-2540
;  INFORMATION FOR SEQ ID NO:  30:
;   SEQUENCE CHARACTERISTICS:
```

```
;       LENGTH:  936 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Melanocarpus albomyces
;       STRAIN:  ALKO4237
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  33..115
;       OTHER INFORMATION:  /codon_start= 33
;       OTHER INFORMATION:  /product= "20K-cellulase"
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  187..435
;       OTHER INFORMATION:  /product= "20K-cellulase"
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  506..881
;       OTHER INFORMATION:  /product= "20K-cellulase"
US-08-841-636A-30

  Query Match             30.1%;  Score 353.6;  DB 7;  Length 936;
  Best Local Similarity   72.8%;  Pred. No. 5.4e-57;
  Matches  505;  Conservative    0;  Mismatches  119;  Indels   70;  Gaps    1;

Qy     145 CGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTTGGCCCGGGAAGGCCGCCGTCAGCC 204
           | |||||||||||||||||||||||||||||||||||  |||| ||| |||| | |||| | ||
Db     183 CCAGATACTGGGACTGCTGCAAGCCGTCGTGCGGCTGGCGCGGAAAGGGCCCCGTGAACC 242

Qy     205 AACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCCTGTCCGACTTCAATGTCCAGTCGG 264
           | || |||||| ||||||| ||||||||||||||| |   ||||||| ||| |   ||||
Db     243 AGCCCGTCTACTCGTGCGACGCCAACTTCCAGCGCATCCACGACTTCGATGCCGTCTCGG 302

Qy     265 GCTGCAACGGCGGCTCGGCCTACTCCTGCGCCGACCAGACTCCCTGGGCGGTGAACGACA 324
           ||||| | |||||| | |||| ||| |||||||||||| |  |||||||  | || ||||
Db     303 GCTGCGAGGGCGGCCCCCGCCTTCTCGTGCGCCGACCACAGCCCCTGGGCCATTAATGACA 362

Qy     325 ATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCGGCGGGTCCGAATCCTCGTGGTGCT 384
           | ||| | |||||||||| |||||     ||  ||||  | ||||    ||||||||||
Db     363 ACCTCTCGTACGGCTTCGCGGCGACTGCACTCAGCGGCCAGACCGAGGAGTCGTGGTGCT 422

Qy     385 GCGCCTGCTACGCG---------------------------------------------- 398
           | ||||||||||||
Db     423 GTGCCTGCTACGCGTGAGTGTGCTTGGGCCCAACGTCGGTGATTCCGGAGTTCAGACCAC 482

Qy     399 ----------------------CTCACCTTCACTTCCGGTCCCGTCGCCGGCAAGACA 434
                                  || ||||| || || |||||||| ||||||||||
Db     483 TGACCCAGCGACCCGCTCGCCAGTCTGACCTTTACATCGGGTCCCGTGGCCGGCAAGACC 542

Qy     435 ATGGTGGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACCAGTTCGATATCGCC 494
           ||||| || ||||| || ||||| |||||||||| || || ||||| |||||  ||  |
Db     543 ATGGTCGTCCAGTCGACCAGCACGGGCGGCGACCTCGGCAGCAACCACTTCGACCTCAAC 602

Qy     495 ATGCCCGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGTTCGGCGGCCTCCCC 554
           || |||||||||||||| ||| ||||| |||||||||   | ||||||||||||||||
Db     603 ATCCCCGGCGGCGGCGTCGGCCTCTTCGACGGCTGCACTCCCCAGTTCGGCGGCCTCCCG 662

Qy     555 GGCGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCTTCCCCGCGCCGCTC 614
           ||||| |  |||||||||| |||||||| |  ||||||| || ||||||| |||||||
Db     663 GGCGCACGGTACGGCGGCATCTCGTCGCGCCAGGAGTGCGACTCGTTCCCCGAGCCGCTC 722

Qy     615 AAGCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACAACCCGACGTTCACG 674
           |||||||||||||||||| || |||||||||||||||||||||||| | || ||
Db     723 AAGCCCGGCTGCCAGTGGCGCTTCGACTGGTTCCAGAACGCCGACAACCCGTCCTTTACC 782

Qy     675 TTCCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCTGCAAGCGCAACGAC 734
           ||| ||| ||| |||||||||| ||| | || || ||  ||||||||| |||| |||||
Db     783 TTCGAGCGGGTCCAGTGCCCCGAGGAGCTGGTCGCTCGGACCGGCTGCAGGCGCCACGAC 842

Qy     735 GACTCCAGCTTCCCCGTCTTCACCCCCCCAAGCG 768
           |||  | ||||| |||||||||   |||| ||||
Db     843 GACGGCGGCTTCGCCGTCTTCAAGGCCCCCAGCG 876


RESULT 3
US-09-735-787-3
; Sequence 3, Application US/09735787
; Patent No. US20010036910A1
;   GENERAL INFORMATION:
;       APPLICANT: Rasmussen, Grethe
;                  Mikkelsen, Jan Moller
;                  Schulein, Martin
;                  Patkar, Shankant A.
;                  Hagen, Fred
;       TITLE OF INVENTION: A Cellulase Preparation Comprising an
;                           Endoglucanase Enzyme
;       NUMBER OF SEQUENCES: 33
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: No. US20010036910A1o No. US20010036910A1disk of No. US200100
;            STREET: 405 Lexington Avenue, 64th Floor
;            CITY: New York
;            STATE: New York
;            COUNTRY: United States of America
;            ZIP: 10174-6401
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/735,787
;            FILING DATE: 13-Dec-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/189,028
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Lambiris, Elias J.
;            REGISTRATION NUMBER: 33,728
;            REFERENCE/DOCKET NUMBER: 3469.214-US
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 212-867-0123
;            TELEFAX: 212-878-9655
;   INFORMATION FOR SEQ ID NO: 3:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1473 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       ORIGINAL SOURCE:
;            ORGANISM: Fusarium oxysporum
;            STRAIN: DSM 2672
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  97..1224
;       SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-735-787-3

  Query Match             26.8%;  Score 315.2;  DB 10;  Length 1473;
  Best Local Similarity   68.5%;  Pred. No. 7.3e-50;
  Matches  466;  Conservative    0;  Mismatches  208;  Indels    6;  Gaps    2;

Qy      83 CACAACCCTGGCCGCTGCACTTCCTCTGGTCGCCTCCGCGGCCAGTGGCAGTGGCCAGTC 142
             |||    || |||   ||    ||||| | ||     || ||   ||| || || || ||
Db     108 CACTCTTCTCGCCCTGGCCGGCCCTCTCGCCGTGAGTGCTGCTTCTGGAAGCGGTCACTC 167

Qy     143 CACGAGATACTGGGACTGCTGCAAGCCGTCGTGCGCTTGGCCCGGGAAGGCCGCCGTCAG 202
             ||  |||||||||| |||||||||| || ||| |||||  ||| ||||| || ||||
Db     168 TACTCGATACTGGGATTGCTGCAAGCCTTCTTGCTCTTGGAGCGGAAAGGCTGCTGTCAA 227

Qy     203 CCAACCGGTCTACGCGTGCGATGCCAACTTCCAGCGCCTGTCCGACTTCAATGTCCAGTC 262
```

```
          |   || |  |   | || |||   |||  | | | | | ||| ||  |||||
Db    228 CGCCCCTGCTTTAACTTGTGATAAGAACGACAACCCCATTTCCAACACCAATGCTGTCAA 287

Qy    263 GGGCTGCAA---CGGCGGCTCGGCCTACTCCTGCGCCGACCAGACTCCCTGGGCGGTGAA 319
          || ||  |    || || || || ||  | ||| || || |  ||||||||| || ||
Db    288 CGGTTGTGAGGGTGGTGGTTCTGCTTATGCTTGCACCAACTACTCTCCCTGGGCTGTCAA 347

Qy    320 CGACAATCTCGCCTACGGCTTCGCCGCGACGAGCATCGCCGGCGGGTCCGAATCCTCGTG 379
          |||  | || |||||||| ||||| || || |  ||| |||| || |||||  ||   ||
Db    348 CGATGAGCTTGCCTACGGTTTCGCTGCTACCAAGATCTCCGGTGGCTCCGAGGCCAGCTG 407

Qy    380 GTGCTGCGCCTGCTACGCGCTCACCTTCACTTCCGGTCCCGTCGCCGGCAAGACAATGGT 439
          |||||| || ||||| ||  | ||||||||  | || ||||||   |||||||  ||| |
Db    408 GTGCTGTGCTTGCTATGCTTTGACCTTCACCACTGGCCCCGTCAAGGGCAAGAAGATGAT 467

Qy    440 GGTGCAGTCAACGAGCACTGGCGGCGACCTGGGAAGTAACCAGTTCGATATCGCCATGCC 499
          || ||||| || | |||||| || || || ||    ||||| |||||| ||   |||||
Db    468 CGTCCAGTCCACCAACACTGGAGGTGATCTCGGCGACAACCACTTCGATCTCATGATGCC 527

Qy    500 CGGCGGCGGCGTGGGCATCTTCAACGGCTGCAGCTCGCAGTTCGGC---GGCCTCCCCGG 556
          |||||| || || || |||||| ||||||||| |||  ||||||||   |  |||  |||
Db    528 CGGCGGTGGTGTCGGTATCTTCGACGGCTGCACCTCTGAGTTCGGCAAGGCTCTCGGCGG 587

Qy    557 CGCTCAATACGGCGGCATTTCGTCGCGCGACCAGTGCGATTCCTTCCCCGCGCCGCTCAA 616
          || || |||||||| || || || ||   | | || |||  || ||||| ||  |||||
Db    588 TGCCCAGTACGGCGGTATCTCCTCCCGAAGCGAATGTGATAGCTACCCCGAGCTTCTCAA 647

Qy    617 GCCCGGCTGCCAGTGGCGGTTTGACTGGTTCCAGAACGCCGACAACCCGACGTTCACGTT 676
          |  ||| ||||| ||||| || ||||||||| ||||||||||||||    ||||| ||
Db    648 GGACGGTTGCCACTGGCGATTCGACTGGTTCGAGAACGCCGACAACCCTGACTTCACCTT 707

Qy    677 CCAGCAGGTGCAGTGCCCCGCCGAGATCGTTGCCCGCTCCGGCTGCAAGCGCAACGACGA 736
            ||||||| |||||||||   |   || | | |  |   || ||||||||  | |||||
Db    708 TGAGCAGGTTCAGTGCCCCAAGGCTCTCCTCGACATCAGTGGATGCAAGCGTGATGACGA 767

Qy    737 CTCCAGCTTCCCCGTCTTCA 756
          |||||||||||| | |||||
Db    768 CTCCAGCTTCCCTGCCTTCA 787


RESULT 4
US-08-841-636A-36
; Sequence 36, Application US/08841636A
; Patent No. US20020168751A1
;  GENERAL INFORMATION:
;    APPLICANT:  Miettinen-Oinonen, Arja
;    APPLICANT:  Londesborough, John
;    APPLICANT:  Vehmaanper , Jari
;    APPLICANT:  Haakana, Heli
;    APPLICANT:  M ntyl , Arja
;    APPLICANT:  Lantto, Raija
;    APPLICANT:  Elovainio, Minna
;    APPLICANT:  Joutsjoki, Vesa
;    APPLICANT:  Paloheimo, Marja
;    APPLICANT:  Suominen, Pirkko
;    TITLE OF INVENTION:  NOVEL CELLULASES, THE GENES ENCODING THEM AND
;    TITLE OF INVENTION:  USES THEREOF
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/841,636A
;      FILING DATE:  30-APR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/005,335
;      FILING DATE:  17-OCT-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/007,926
;      FILING DATE:  04-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/020,840
;      FILING DATE:  28-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/732,181
;      FILING DATE:  16-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FI96/00550
;      FILING DATE:  17-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Timothy J. Shea, Jr.
;      REGISTRATION NUMBER:  41,306
;      REFERENCE/DOCKET NUMBER:  1716.0510005/MAC/TJS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)371-2600
;      TELEFAX:  (202)371-2540
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  887 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Melanocarpus albomyces
;      STRAIN:  ALKO4237
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  351..455
;      OTHER INFORMATION:  /product= "protein-with-CBD"
US-08-841-636A-36

  Query Match             7.1%;  Score 83.2;  DB 7;  Length 887;
  Best Local Similarity   73.6%;  Pred. No. 4.9e-07;
  Matches  106;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy    817 GCCAGACGTCTCCCGGCGGCGGCAGTGGCTGCACGTCTCAGAAGTGGGCTCAGTGCGGTG 876
          || |  ||     |||||||||| | ||||||||    ||||||||||  |||||||| |
Db    313 GCAACCCGGGCAACGGCGGCGGCGGCGGCTGCACCGTCCAGAAGTGGGGCCAGTGCGGCG 372

Qy    877 GCATCGGCTTCAGCGGATGCACCACCTGTGTCTCTGGCACCACCTGCCAGAAGTTGAACG 936
          ||||||||| |   || |||||||||||     | ||| | |||||||  |     |||||
Db    373 GCATCGGCTACTCGGGCTGCACCACCTGCAAGGCCGGCTCGACCTGCCCGGCCCAGAACG 432

Qy    937 ACTACTACTCGCAGTGCCTCTAAA 960
          | ||||||||||||||||| ||||
Db    433 AGTACTACTCGCAGTGCCTGTAAA 456


RESULT 5
US-09-764-847-20
; Sequence 20, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009
;  CURRENT APPLICATION NUMBER: US/09/764,847
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2003
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 682
;   TYPE: DNA
;   ORGANISM: Homo sapiens
```

```
US-09-764-847-20

  Query Match             5.6%;  Score 66;  DB 10;  Length 682;
  Best Local Similarity   83.3%;  Pred. No. 0.00071;
  Matches   75;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy   1085 ACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAAAAAA 1144
          || ||  | | || | | || | || |||  | | ||||||||||||||||||||||||
Db    563 ACGCCAGCCTGGGCGACAGAGTGAGACCCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAA 622

Qy   1145 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
          ||||||||||||||||||||||||||||||
Db    623 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 652


RESULT 6
US-09-925-299-16
; Sequence 16, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
;   LENGTH: 371
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (350)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (360)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-16

  Query Match             5.6%;  Score 65.6;  DB 10;  Length 371;
  Best Local Similarity   60.8%;  Pred. No. 0.00075;
  Matches  107;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy    999 CCGTCAACTTCGTCATGCATATTTTTTGAGCGCTCAATACATACATAACCTTCGATTCTT 1058
          ||  |  ||   |  || | ||   | |   |  | | ||||   |  | ||  || |||
Db    169 CCAACCCCTATATATTGTAAATAGATGGGCTGGGCTAAACATTGTTGCCGTTTCATACTT 228

Qy   1059 GTACATAGCACGCCGGTACACATCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTT 1118
           |||  |    ||  ||||||    |   | ||| |    |   ||  |
Db    229 CTACCAACTCAGCTTTTACACAATAAAGCTCTACTGTCTCTGAAAAAAAAAAAAAAAAA 288

Qy   1119 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    289 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 344


RESULT 7
US-09-764-846-41
; Sequence 41, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ12
;  CURRENT APPLICATION NUMBER: US/09/764,846
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 348
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
;   LENGTH: 1341
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-846-41

  Query Match             5.5%;  Score 64.6;  DB 10;  Length 1341;
  Best Local Similarity   83.9%;  Pred. No. 0.0014;
  Matches   73;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy   1088 CCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1147
          | ||| | ||   | || | || |||  | | ||||||||||||||||||||||||||||
Db   1217 CAGACCTGGGCAACAGAGTGAGACCCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1276

Qy   1148 AAAAAAAAAAAAAAAAAAAAAAAAAA 1174
          ||||||||||||||||||||||||||
Db   1277 AAAAAAAAAAAAAAAAAAAAAAAAAA 1303


RESULT 8
US-09-919-580-246/c
; Sequence 246, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
;  APPLICANT: Pyle, Ruth
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.552
;  CURRENT APPLICATION NUMBER: US/09/919,580
;  CURRENT FILING DATE:  2001-07-30
;  NUMBER OF SEQ ID NOS: 934
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 246
;   LENGTH: 95
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 73, 78
;   OTHER INFORMATION: n = A,T,C or G
US-09-919-580-246

  Query Match             5.5%;  Score 64;  DB 10;  Length 95;
  Best Local Similarity   97.0%;  Pred. No. 0.0012;
  Matches   64;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy   1109 GGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1168
          ||||| |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     83 GGCCCNTTTTNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 24

Qy   1169 AAAAAA 1174
          ||||||
Db     23 AAAAAA 18


RESULT 9
US-09-745-763-69
; Sequence 69, Application US/09745763
; Patent No. US20020065394A1
;   GENERAL INFORMATION:
;        APPLICANT: Jacobs, Kenneth
;                   McCoy, John M.
;                   LaVallie, Edward R.
;                   Collins-Racie, Lisa A.
;                   Evans, Cheryl
;                   Merberg, David
;                   Treacy, Maurice
;                   Spaulding, Vikki
;        TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
;                            ENCODING THEM
```

```
;       NUMBER OF SEQUENCES: 219
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genetics Institute, Inc.
;            STREET: 87 CambridgePark Drive
;            CITY: Cambridge
;            STATE: MA
;            COUNTRY: U.S.A.
;            ZIP: 02140
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/745,763
;            FILING DATE: 18-Jun-2000
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Sprunger, Suzanne A.
;            REGISTRATION NUMBER: 41,323
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617) 498-8284
;            TELEFAX: (617) 876-5851
;   INFORMATION FOR SEQ ID NO: 69:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1908 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 69:
US-09-745-763-69

  Query Match             5.4%;  Score 63.8;  DB 10;  Length 1908;
  Best Local Similarity   69.9%;  Pred. No. 0.0021;
  Matches   86;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy    1052 GATTCTTGTACATAGCACGCCGGTACACATCTCACACCGACTTTGGGGGCGGAATCAGGC 1111
           |||   |||    |||  |   |    |    ||| ||| | | || ||| || | || |
Db    1778 GATGGCTGTGGTGAGCCGGGATGGCGCCGCTGCACTCCGGCCTGGGCGGCTGAGTGAGAC 1837

Qy    1112 CCGTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1171
             |  |     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1838 TCCATCCCCGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1897

Qy    1172 AAA 1174
           |||
Db    1898 AAA 1900


RESULT 10
US-09-925-301-29
; Sequence 29, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA106
;  CURRENT APPLICATION NUMBER: US/09/925,301
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05882
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1694
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
;   LENGTH: 917
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-301-29

  Query Match             5.4%;  Score 63.6;  DB 10;  Length 917;
  Best Local Similarity   83.7%;  Pred. No. 0.0021;
  Matches   72;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy    1089 CGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1148
           | || ||| |||   || ||| |   ||   ||||||||||||||||||||||||||||
Db     814 CTACCTTGTGGGAATAAACAGACAAATTAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 873

Qy    1149 AAAAAAAAAAAAAAAAAAAAAAAAA 1174
           |||||||||||||||||||||||||
Db     874 AAAAAAAAAAAAAAAAAAAAAAAAA 899


RESULT 11
US-09-822-830A-588
; Sequence 588, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
;  APPLICANT: Genetics Institute, Inc.
;  APPLICANT:  Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Howes, Steven H.
;  APPLICANT:  Resnick, Richard J.
;  APPLICANT:  Gulukota, Kamalakar
;  APPLICANT:  Graham, James R.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
;  FILE REFERENCE: GIN 6402
;  CURRENT APPLICATION NUMBER: US/09/822,830A
;  CURRENT FILING DATE:  2001-03-29
;  PRIOR APPLICATION NUMBER: 60/195,604
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 631
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 588
;   LENGTH: 1377
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-822-830A-588

  Query Match             5.4%;  Score 63.4;  DB 10;  Length 1377;
  Best Local Similarity   82.0%;  Pred. No. 0.0024;
  Matches   73;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy    1086 CACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAA 1145
           | | || || ||| |   || |||    |  || ||||||||||||||||||||||||||
Db    1262 CCCTGATTTGGGGAGAATAAACAGATATGAGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAA 1321

Qy    1146 AAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
           ||||||||||||||||||||||||||||
Db    1322 AAAAAAAAAAAAAAAAAAAAAAAAAAAA 1350


RESULT 12
US-09-960-352-14757/c
; Sequence 14757, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14757
;   LENGTH: 355
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 63-LIB3058-014-Q1-K1-H4
```

```
US-09-960-352-14757

  Query Match             5.4%;  Score 63.2;  DB 10;  Length 355;
  Best Local Similarity   95.6%;  Pred. No. 0.0021;
  Matches   65;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy   1107 CAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1166
          | | ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    147 CGGCCCCCTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 88

Qy   1167 AAAAAAAA 1174
          ||||||||
Db     87 AAAAAAAA 80


RESULT 13
US-09-764-877-3172/c
; Sequence 3172, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC005
;  CURRENT APPLICATION NUMBER: US/09/764,877
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 4031
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3172
;   LENGTH: 1545
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-877-3172

  Query Match             5.4%;  Score 63;  DB 10;  Length 1545;
  Best Local Similarity   78.9%;  Pred. No. 0.0029;
  Matches   75;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy   1079 CATCTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAA 1138
          |||  ||| ||  | | || | | ||   ||   |  | | ||||||||||||||||||
Db    843 CATTGCACTCCAGCCTGGGTGACAGAGCAAGATTCCGTCTCAAAAAAAAAAAAAAAAAAA 784

Qy   1139 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1173
          ||||||||||||||||||||||||||||||||||||
Db    783 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 749


RESULT 14
US-09-960-352-14316/c
; Sequence 14316, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14316
;   LENGTH: 391
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 61-LIB3058-052-Q1-K1-H10
US-09-960-352-14316

  Query Match             5.3%;  Score 62.6;  DB 10;  Length 391;
  Best Local Similarity   88.3%;  Pred. No. 0.0027;
  Matches   68;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy   1098 GGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1157
          ||||  |||  || |||    ||||||||||||||||||||||||||||||||||||||||||
Db    137 GGGCCCAATGTGGGCCGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 78

Qy   1158 AAAAAAAAAAAAAAAA 1174
          ||||||||||||||||
Db     77 AAAAAAAAAAAAAAAA 61


RESULT 15
US-09-764-864-197
; Sequence 197, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 197
;   LENGTH: 1817
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-864-197

  Query Match             5.3%;  Score 62.6;  DB 10;  Length 1817;
  Best Local Similarity   79.6%;  Pred. No. 0.0035;
  Matches   74;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy   1082 CTCACACCGACTTTGGGGGCGGAATCAGGCCCGTTTTAAAAAAAAAAAAAAAAAAAAAAA 1141
          ||| | | |||  |   | | ||||  |  | |    ||||||||||||||||||||||
Db   1718 CTCTCCCTGACAATAAAGTCTGAATTTGTTCTGCCAAAAAAAAAAAAAAAAAAAAAAAAA 1777

Qy   1142 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
          ||||||||||||||||||||||||||||||||||
Db   1778 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1810


Search completed: February 27, 2003, 18:37:26
Job time : 147 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 24, 2003, 11:42:49 ; Search time 85 Seconds
(without alignments)
468.729 Million cell updates/sec

Title: US-10-007-521-12
Perfect score: 1659
Sequence: 1 MRSTPVLRTTLAAALPLVAS.....TTCVSGTTCQKLNDYYSQCL 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1659	100.0	299	17	AAW04928	Cellulytic enzyme
2	1659	100.0	299	19	AAW63624	Monocomponent endo
3	1178	71.0	305	19	AAW44854	Humicola insolens
4	1178	71.0	305	19	AAW41929	Humicola insolens
5	1159	69.9	305	12	AAR15271	Humicola insolens
6	1159	69.9	305	13	AAR25525	Humicola insolens
7	1159	69.9	305	13	AAR25464	Endoglucanase #1.
8	1159	69.9	305	13	AAR25428	Cellulase containe
9	1159	69.9	305	13	AAR27968	Endoglucanase enzy
10	1159	69.9	305	13	AAR28295	Sequence of ~ 43 k
11	1159	69.9	305	14	AAR37150	Dye transfer inhib
12	1159	69.9	305	14	AAR42063	Endoglucanase enzy
13	1159	69.9	305	16	AAY67388	H. insolens endogl
14	1159	69.9	305	16	AAW01502	43 kD endoglucanas
15	1159	69.9	305	17	AAR88471	Alkaline endogluca
16	1159	69.9	305	19	AAW46616	Humicola insolens
17	1159	69.9	305	19	AAW44266	Humicola insolens
18	1159	69.9	305	21	AAB03660	Endoglucanase amin
19	1146	69.1	305	17	AAG78352	Humicola insolens
20	1145	69.0	305	17	AAG78356	Humicola insolens
21	1143	68.9	305	17	AAG78359	H. insolens mutant
22	1142.5	68.9	306	19	AAW44269	Hybrid DNA protein
23	1142	68.8	305	13	AAR28300	43kD endoglucanase
24	1142	68.8	305	17	AAG78355	Humicola insolens
25	1142	68.8	305	17	AAG78358	Humicola insolens
26	1138	68.6	305	17	AAG78353	Humicola insolens
27	1137	68.5	305	17	AAG78354	Humicola insolens
28	1137	68.5	305	17	AAG78357	Humicola insolens
29	1137	68.5	305	17	AAG78360	Humicola insolens
30	1134	68.4	305	13	AAR28818	H. insolens cellul
31	1127	67.9	286	19	AAW57420	Humicola insolens
32	1122	67.6	200	19	AAW53968	Thielavia terrestr
33	1120	67.5	284	15	ABB04137	Mutant 43kD endogl
34	1119	67.5	284	15	ABB04129	Mutant 43kD endogl
35	1117	67.3	200	19	AAW53979	Thielavia terrestr
36	1114	67.1	284	15	ABB04141	Humicola insolens
37	1113	67.1	284	15	ABB04140	Mutant 43kD endogl
38	1112	67.0	284	15	ABB04128	Humicola insolens
39	1112	67.0	284	15	ABB04130	Mutant 43kD endogl
40	1110	66.9	305	17	AAG78361	Humicola insolens
41	1108	66.8	284	15	ABB04131	Mutant 43kD endogl
42	1108	66.8	284	15	ABB04135	Mutant 43kD endogl
43	1107.5	66.8	306	19	AAW44270	Hybrid DNA protein
44	1107	66.7	284	15	ABB04132	Mutant 43kD endogl
45	1106.5	66.7	267	15	ABB04136	Mutant 43kD endogl

ALIGNMENTS

RESULT 1
AAW04928
ID AAW04928 standard; Protein; 299 AA.
XX
AC AAW04928;
XX
DT 20-MAY-1997 (first entry)
XX
DE Cellulytic enzyme #4 of the invention.
XX
KW Cellulytic enzyme; endoglucanase; hydrolysis; cellulose; microorganism;
KW plant cellulase; catalytic region; textile; backstaining; bio-polishing;
KW stone-washing; cellulosic fabric; colour clarification; defibration;
KW cell wall degradation; paper pulp; debarking; fibre modification;
KW enzymatic de-inking; drainage improvement.
XX
OS Thielavia terrestris.
XX
PN WO9629397-A1.
XX
PD 26-SEP-1996.
XX
PF 18-MAR-1996; 96WO-DK00105.
XX
PR 12-FEB-1996; 96DK-0000137.
PR 17-MAR-1995; 95DK-0000272.
PR 08-AUG-1995; 95DK-0000885.
PR 08-AUG-1995; 95DK-0000886.
PR 08-AUG-1995; 95DK-0000887.
PR 08-AUG-1995; 95DK-0000888.
XX
PA (NOVO) NOVO-NORDISK AS.
XX

PI Andersen LN, Ihara M, Kauppinen MS, Lange L, Lassen SF;
PI Nielsen RI, Schuelein M, Takagi S;
XX
DR WPI; 1996-443173/44.
DR N-PSDB; AAT39050.
XX
PT New endo:glucanase enzyme preparations - contg. conserved catalytic
PT regions, useful for treating fabrics, textiles, plant material or
PT paper pulp
XX
PS Claim 72; Page 161-163; 316pp; English.
XX
CC AAW04925-W04932 represent the enzymes of the invention. These enzymes
CC possess cellulytic (particularly endoglucanase) activity. Cellulytic
CC enzymes are involved in the the hydrolysis of cellulose, and are
CC synthesised by a large number of microorganisms and plants. The enzymes
CC of the invention containing the conserved catalytic regions (such as
CC AAW04913) exhibit improved performance, e.g. 50 times higher
CC performance, compared to multiple domain enzymes. The enzymes can be
CC used for the treatment of fabrics or textiles, preferably for preventing
CC backstaining, for bio-polishing or for stone-washing cellulosic fabric.
CC They can also be used to provide colour clarification for laundry. The
CC enzymes can also be used for the degradation or modification of plant
CC material, such as cell walls. They can also be used in the treatment of
CC paper pulp preferably for debarking, defibration, fibre modification,
CC enzymatic de-inking or drainage improvement.
XX
SQ Sequence 299 AA;

Query Match 100.0%; Score 1659; DB 17; Length 299;
Best Local Similarity 100.0%; Pred. No. 2.5e-119;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60

Qy      61 LSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGP 120

Qy     121 VAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCDS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 VAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCDS 180

Qy     181 FPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 FPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNG 240

Qy     241 GTGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
```

RESULT 2
AAW63624
ID AAW63624 standard; Protein; 299 AA.
XX
AC AAW63624;
XX
DT 21-SEP-1998 (first entry)
XX
DE Monocomponent endoglucanase enzyme.
XX
KW Monocomponent endoglucanase; cellulolytic enzyme; garment; abrasion;
KW abraded looking jeans; fungus; Thielavia terrestris; stone-washed.
XX
OS Thielavia terrestris.
XX
PN EP843041-A1.
XX
PD 20-MAY-1998.
XX
PF 30-MAY-1997; 97EP-0610021.
XX
PR 13-NOV-1996; 96DK-0001276.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Kalum L, Lund H;
XX
DR WPI; 1998-263308/24.
DR N-PSDB; AAV39096.
XX
PT Pair of jeans - with abraded look
XX
PS Claim 16; Pages 10-12; 16pp; English.
XX
CC This represents a monocomponent endoglucanase. This is a cellulolytic
CC enzyme belonging to the fungal species Thielavia terrestris. This can be
CC used in the process of invention of providing a pair of jeans made from
CC dyed twill fabric that has localised variations in the colour density of
CC the fabric. The process involves providing the jeans with a stone-washed
CC or abraded look, where the variation corresponds to a delta remission
CC value (Delta R) higher than 11, and the reflection of a first area of
CC the jeans fabric is less than 12 percent. The process comprises selecting
CC the desired textile cutting pattern for jeans garment, positioning the
CC pattern onto newly manufactured dyed twill fabric, cutting the jeans
CC garment parts, sowing the pair of jeans, optionally subjecting the pair
CC of jeans to a desizing treatment, and subjecting the pair of jeans to an
CC abrasion treatment with an efficient amount of a cellulolytic enzyme in
CC an aqueous medium essentially free of bleaching chemicals. The abraded
CC look near stitching is different from the abraded look far from
CC stitching.
XX
SQ Sequence 299 AA;

Query Match 100.0%; Score 1659; DB 19; Length 299;
Best Local Similarity 100.0%; Pred. No. 2.5e-119;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60

Qy      61 LSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGP 120

Qy     121 VAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCDS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 VAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCDS 180

Qy     181 FPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 FPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNG 240

Qy     241 GTGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
```

RESULT 3
AAW44854
ID AAW44854 standard; Protein; 305 AA.
XX
AC AAW44854;
XX
DT 31-JUL-1998 (first entry)
XX
DE Humicola insolens cellulase NCE4 protein.
XX
KW Humicola insolens; NCE1; NCE2; NCE4; cellulase; expression vector;
KW promoter; signal sequence; terminator; amylase; lipase; protease;
KW phytase.
XX

```
OS   Humicola insolens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= signal
FT   Protein         22..305
FT                   /label= Cellulase_NCE4
XX
PN   WO9803667-A1.
XX
PD   29-JAN-1998.
XX
PF   24-JUL-1997;   97WO-JP02560.
XX
PR   24-JUL-1996;   96JP-0195070.
XX
PA   (MEIJ ) MEIJI SEIKA KAISHA LTD.
XX
PI   Aoyagi K,  Hamaya T,  Koga J,  Kono T,  Moriya T,  Murakami T;
PI   Murashima K, Sumida N,  Watanabe M;
XX
DR   WPI; 1998-120786/11.
DR   N-PSDB; AAV19378.
XX
PT   Mass production of proteins and peptides in Humicola species - using
PT   expression vector containing the promoter, signal sequence and/or
PT   terminator from the Humicola insolens NCE1 or NCE2 gene
XX
PS   Example A6; Page 46-48; 63pp; Japanese.
XX
CC   The present sequence represents the Humicola insolens cellulase NCE4
CC   protein from the present invention. The present invention describes a
CC   method for the mass production of proteins and peptides in Humicola
CC   species, especially in Humicola insolens, using an expression vector
CC   which comprises the promoter, signal sequence and/or terminator
CC   regulatory sequences from the NCE1 or NCE2 gene of H. insolens. These
CC   are available in the plasmids pM3-1 (Escherichia coli JM109/pM3-1,
CC   FERM BP-5971) (for NCE1) and pM14-1 (E. coli JM109/pM14-1, FERM BP-5972)
CC   (for NCE2). The vector also contains a marker gene such as an antibiotic
CC   resistance gene (e.g. the destomycin resistance gene from Streptomyces
CC   rimofaciens). Proteins which can be expressed using this system include
CC   cellulase, amylase, lipase, protease, phytase and other enzymes.
CC   Specific expression vectors of the invention are pMKD01 (for Humicola
CC   NCE3 cellulase gene), pEGD01 (for Humicola NCE4 cellulase gene) and
CC   pIED02 (for Humicola NCE4 cellulase gene). The expression system allows
CC   the efficient production of proteins and peptides in a Humicola host.
CC   Using the expression system high amounts of protein (>4.5 g/l) can be
CC   obtained.
XX
SQ   Sequence   305 AA;

  Query Match             71.0%;  Score 1178;  DB 19;  Length 305;
  Best Local Similarity   68.1%;  Pred. No. 1.8e-82;
  Matches  209;  Conservative   44;  Mismatches   44;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:||: : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLRSAVVAALPVLALAA--DGKSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           |:||: :|||  || |||||||||||||: |:|||||||| :|: |||||| ||||||
Db      59 LTDFDAKSGCEPGGVAYSCADQTPWAVNDDFAFGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :   | ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|

Db     239 SSPVGQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 4
AAW41929
ID   AAW41929 standard; Protein; 305 AA.
XX
AC   AAW41929;
XX
DT   26-JUN-1998  (first entry)
XX
DE   Humicola insolens endoglucanase cellulase NCE4.
XX
KW   Endoglucanase; cellulase; NCE4; fluff removal; weight reduction;
KW   denier reduction; bleaching; denim dyed fibre.
XX
OS   Humicola insolens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /label= sig_peptide
FT   Peptide         23..305
FT                   /label= mat_peptide
XX
PN   WO9803640-A1.
XX
PD   29-JAN-1998.
XX
PF   24-JUL-1997;   97WO-JP02561.
XX
PR   24-JUL-1996;   96JP-0194974.
XX
PA   (MEIJ ) MEIJI SEIKA KAISHA LTD.
XX
PI   Aoyagi K,  Hamaya T,  Koga J,  Kono T,  Moriya T,  Murakami T;
PI   Murashima K,  Sumida N;
XX
DR   WPI; 1998-120765/11.
DR   N-PSDB; AAV13840.
XX
PT   Endoglucanase cellulase NCE4 from Humicola insolens - for treatment
PT   of cellulose fibres for fluff elimination and weight reduction
XX
PS   Claim 1; Pages 33-36; 43pp; Japanese.
XX
CC   The present sequence is the Humicola insolens
CC   endoglucanase cellulase NCE4, which can be used to treat cellulose
CC   fibres for fluff removal, weight (denier) reduction and bleaching,
CC   especially of denim dyed fibres.
XX
SQ   Sequence   305 AA;

  Query Match             71.0%;  Score 1178;  DB 19;  Length 305;
  Best Local Similarity   68.1%;  Pred. No. 1.8e-82;
  Matches  209;  Conservative   44;  Mismatches   44;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:||: : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLRSAVVAALPVLALAA--DGKSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           |:||: :|||  || |||||||||||||: |:|||||||| :|: |||||| ||||||
Db      59 LTDFDAKSGCEPGGVAYSCADQTPWAVNDDFAFGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
```

```
          ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db    179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy    239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
          :   | ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db    239 SSPVGQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy    293 DYYSQCL 299
          |:| |||
Db    299 DWYHQCL 305


RESULT 5
AAR15271
ID   AAR15271 standard; Protein; 305 AA.
XX
AC   AAR15271;
XX
DT   18-FEB-1992  (first entry)
XX
DE   Humicola insolens DSM 1800 endoglucanase.
XX
KW   cellulase; cellulose.
XX
OS   Humicola insolens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= signal
FT   Protein         22..305
FT                   /label= endoglucanase
XX
PN   WO9117243-A.
XX
PD   14-NOV-1991.
XX
PF   08-MAY-1991;   91WO-DK00123.
XX
PR   22-APR-1991;   91DK-0000736.
PR   09-MAY-1990;   90DK-0001159.
XX
PA   (NOVO ) NOVO NORDISK A/S.
XX
PI   Rasmussen G,  Mikkelsen JM,  Schulein M,  Patkar SA,  Hagen F;
PI   Hjort CM,  Hastrup S;
XX
DR   WPI; 1991-353765/48.
DR   N-PSDB; AAQ14856.
XX
PT   Cellulase prepn. comprising endoglucanase enzyme - used in
PT   detergents for cellulose-contg. fabrics or to improve drainage of
PT   paper pulp
XX
PS   Claim 6; Page 48; 67pp; English.
XX
CC   The cellulase isolated from H.insolens has mol.wt. ca. 43kD by
CC   SDS-PAGE and an isoelectric point of ca. 5.1. The enzyme can be
CC   used in detergents for softening, soil removal and colour
CC   clarification. It can reduce the harshness of cellulose-contg.
CC   fabrics.
XX
SQ   Sequence   305 AA;

  Query Match             69.9%;  Score 1159;  DB 12;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 5.2e-81;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy      1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
          |||:|:| : : ||||::| ||   |:|||||||||||| |  || |:|||::|:|||||
Db      1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy     61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
          ::||: :|||  || ||||||||||||||: | ||||||||| :|: |||||| ||||||
```

```
Db     59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy    120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
          ||||| |||||||||||||| ||: :||||||||:||: ||||||| :|||||||::||
Db    119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy    180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
          ||  ||||| |||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db    179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy    239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
          :     ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db    239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy    293 DYYSQCL 299
          |:| |||
Db    299 DWYHQCL 305


RESULT 6
AAR25525
ID   AAR25525 standard; Protein; 305 AA.
XX
AC   AAR25525;
XX
DT   11-JAN-1993  (first entry)
XX
DE   Humicola insolens DSM 1800 endoglucanase.
XX
KW   Detergent; washing powder; cellulase; softening clay; C14CMC-method;
KW   flocculation; radio-labelled carboxymethylcellulase.
XX
OS   Humicola insolens DSM 1800.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= signal
FT   Protein         22..305
FT                   /label= endoglucanase
XX
PN   EP495258-A.
XX
PD   22-JUL-1992.
XX
PF   06-NOV-1991;   91EP-0202880.
XX
PR   16-JAN-1991;   91EP-0870006.
PR   06-NOV-1991;   91EP-0202880.
PR   06-NOV-1991;   91EP-0202879.
XX
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Baeck AC,  Busch A,  Convents AC;
XX
DR   WPI; 1992-243163/30.
DR   N-PSDB; AAQ26405.
XX
PT   Compact, granular detergent compsns. - contain high activity
PT   cellulase and softening clay to provide synergistic effect in
PT   softening performance
XX
PS   Claim 5; Page 29-30; 38pp; English.
XX
CC   The endoglucanase isolated from H.insolens DSM 1800 has Mr 43,000
CC   and high cellulase activity (i.e. removes a minimum of 10% of
CC   radiolabelled carboxymethylcellulose (C14CMC)). The enzyme can be
CC   used in detergent compositions along with a surface active agent,
CC   a builder system and a softening clay. See also AAQ26407.
XX
SQ   Sequence   305 AA;

  Query Match             69.9%;  Score 1159;  DB 13;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 5.2e-81;
```

```
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :||| || ||||||||||||: | ||||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :||||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| |||||:||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :     ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 7
AAR25464
ID   AAR25464 standard; Protein; 305 AA.
XX
AC   AAR25464;
XX
DT   07-JAN-1993  (first entry)
XX
DE   Endoglucanase #1.
XX
KW   CMC-endoase; 43 kD cellulase; monoclonal antibody.
XX
OS   Humicola insolens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= Signal_peptide
FT   Protein         22..305
FT                   /label= Mature_protein
XX
PN   EP495257-A.
XX
PD   22-JUL-1992.
XX
PF   06-NOV-1991;   91EP-0202879.
XX
PR   16-JAN-1991;   91EP-0870006.
PR   06-NOV-1991;   91EP-0202880.
PR   06-NOV-1991;   91EP-0202879.
XX
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Baeck AC,  Busch A,  Ceulemans RAA;
XX
DR   WPI; 1992-243163/30.
DR   N-PSDB; AAQ26380.
XX
PT   Compact, granular detergent compsns. - contain high activity
PT   cellulase and softening clay to provide synergistic effect in
PT   softening performance
XX
PS   Disclosure; Page 20-21; 29pp; English.
XX
CC   The sequences given in AAR25464 and AAR25466 are endoglucanases which
CC   are immunoreactive with a monoclonal antibody raised against a
CC   partially purified 43 kD cellulase derived from Humicola insolens.
CC   These endoglucanases exhibit a CMC-endoase activity of at least
CC   about 50, pref. at least about 60, inparticular at least about 90 CMC-
CC   endoase units per mg of total protein.  These endoglucanases have
CC   molecular weight of approx. 43 kD.
XX
SQ   Sequence   305 AA;

  Query Match             69.9%;  Score 1159;  DB 13;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 5.2e-81;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :||| || ||||||||||||: | ||||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :||||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| |||||:||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :     ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 8
AAR25428
ID   AAR25428 standard; Protein; 305 AA.
XX
AC   AAR25428;
XX
DT   15-JAN-1993  (first entry)
XX
DE   Cellulase contained in a detergent compsn.
XX
KW   Endoglucanase; immunoreactive; Humicola isolens; cleaning;
KW   softening.
XX
OS   Humicola isolens.
XX
PN   EP495554-A.
XX
PD   22-JUL-1992.
XX
PF   15-JAN-1992;   92EP-0200101.
XX
PR   16-JAN-1991;   91EP-0870006.
PR   06-NOV-1991;   91EP-0202879.
PR   06-NOV-1991;   91EP-0202881.
XX
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Baeck AC,  Boutique J,  Bush A,  Convents AC,  Hargrove RS;
PI   Prettyaj,  Busch A,  Ceulemans RAA;
XX
DR   WPI; 1992-243405/30.
DR   P-PSDB; AAR25428.
XX
PT   Detergent compsn. for improving cleaning and performance - is
PT   composed of quat. ammonium cpd. and high activity cellulose e.g.
PT   homogeneous endo:glucanase
XX
```

```
PS   Claim 10; Page 19; 28pp; English.
XX
CC   The cellulase was used in a detergent compsn. contg. a quaternary
CC   ammonium cpd.  The compsn. gives a detergent with good cleaning
CC   and softening performance due to the synergistic effect of
CC   the ammonium and cellolase components.   The cellulase is a
CC   homogeneous endoglucanase componant which is immunoreactive with an
CC   antibody raised against a cellulase purified from Humicola isolens
CC   DSM 1800.   See also AAR25429.
XX
SQ   Sequence   305 AA;

  Query Match             69.9%;  Score 1159;  DB 13;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 5.2e-81;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:|||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || ||||||||||||||: | |||||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
             ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :    ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 9
AAR27968
ID   AAR27968 standard; Protein; 305 AA.
XX
AC   AAR27968;
XX
DT   09-MAR-1993  (first entry)
XX
DE   Endoglucanase enzyme.
XX
KW   Alkaline cellulase; laundry detergent compositions;
KW   fabric colour maintenance.
XX
OS   Humicola insolens DSM 1800.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "signal peptide"
FT   Peptide         22..305
FT                   /note= "mature peptide"
XX
PN   EP508358-A.
XX
PD   14-OCT-1992.
XX
PF   07-APR-1992;   92EP-0105956.
XX
PR   12-APR-1991;   91EP-0870062.
XX
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Busch A,  MacCorquodale F;
XX
DR   WPI; 1992-341667/42.
DR   N-PSDB; AAQ29934.
XX
PT   Laundry detergent compsns. - contg. alkali cellulase and PVP for
PT   fabric colour maintenance upon laundering
XX
PS   Disclosure; Page 16; 23pp; English.
XX
CC   The sequence is that of an endoglucanase (alkaline cellulase)
CC   from Humicola insolens which is used with a polyvinylpyrrolidone
CC   (PVP) in a laundry detergent compsn. The action of the PVP and the
CC   endoglucanase in fabric colour maintenance upon laundering is superior
CC   to the sum of the individual actions of both these ingredients.
CC   See also AAR27699.
XX
SQ   Sequence   305 AA;

  Query Match             69.9%;  Score 1159;  DB 13;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 5.2e-81;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:|||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || ||||||||||||||: | |||||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
             ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :    ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 10
AAR28295
ID   AAR28295 standard; Protein; 305 AA.
XX
AC   AAR28295;
XX
DT   02-APR-1993  (first entry)
XX
DE   Sequence of ~ 43 kD endogulcanase.
XX
KW   Detergent addative; protease; enzyme; endoglucanase.
XX
OS   Humicola insolens DSM 1800.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= signal
FT   Protein         22..305
XX
PN   WO9218599-A.
XX
PD   29-OCT-1992.
XX
PF   10-APR-1992;   92WO-DK00116.
XX
PR   22-APR-1991;   91DK-0000737.
```

```
XX
PA   (NOVO ) NOVO-NORDISK AS.
XX
PI   Tsuchiya R,  Wagner P;
XX
DR   WPI; 1992-382092/46.
DR   N-PSDB; AAQ30067.
XX
PT   Detergent additive contg. cellulase and specific protease - which
PT   does not degrade the cellulase during storage and clarifies the
PT   colour of dyed cellulosic materials
XX
PS   Disclosure; Page 8-9; 15pp; English.
XX
CC   The ~43 kD endoglucanase derived from Humicola insolens, DSM 1800 is
CC   described in detail in co-pending Danish patent application No. DK
CC   1159/90.  As a detergent additive, it has a higher degree of specificity
CC   then Bacillus lentus serine protease.  The term "higher degree of
CC   specificity" is defined as a protease which conditions degrades human
CC   insulin to fewer components.
XX
SQ   Sequence   305 AA;

  Query Match             69.9%;  Score 1159;  DB 13;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 5.2e-81;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || |||||||||||||: | |||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| ||||||:|||||:|:|:||||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :    ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 11
AAR37150
ID   AAR37150 standard; Protein; 305 AA.
XX
AC   AAR37150;
XX
DT   25-AUG-1993  (first entry)
XX
DE   Dye transfer inhibiting compsn. cellulase.
XX
KW   Detergent; homogeneous endoglucanase component; 43kD cellulase.
XX
OS   Humicola insolens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "signal peptide"
FT   Peptide         22..305
FT                   /note= "mature peptide"
XX
PN   EP540784-A.
XX
PD   12-MAY-1993.
XX
PF   06-NOV-1991;   91EP-0202882.
XX
PR   06-NOV-1991;   91EP-0202882.
XX
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Busch A,  McCorquodale F;
XX
DR   WPI; 1993-153868/19.
DR   N-PSDB; AAQ41732.
XX
PT   Compsn. for inhibiting dye transfer during fabric washing - contains
PT   peroxidase, hydrogen peroxide, substrate and cellulase, esp. endo-
PT   glucanase from Humicola insolens
XX
PS   Disclosure; Page 18-19; 28pp; English.
XX
CC   The sequence is that of a cellulase which is characterised in
CC   that the cellulase provides at least 10% removal of immobilised
CC   radioactive labelled carboxymethyl cellulose according to the C14CMC
CC   method at 25 x 10(-6)% by weight of the cellulase protein in the
CC   test solution. It can be used as part of a compsn. for inhibiting
CC   dye transfer.
XX
SQ   Sequence   305 AA;

  Query Match             69.9%;  Score 1159;  DB 14;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 5.2e-81;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || |||||||||||||: | |||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| ||||||:|||||:|:|:||||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :    ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 12
AAR42063
ID   AAR42063 standard; Protein; 305 AA.
XX
AC   AAR42063;
XX
DT   28-APR-1994  (first entry)
XX
DE   Endoglucanase enzyme.
XX
KW   Detergent; fabric; surfactant; softening clay; cellulase;
KW   Humicola insolens; Fusarium oxysporum; endoglucanase.
XX
OS   Humicola insolens.
XX
```

```
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= sig_peptide
FT   Protein         22..305
FT                   /label= mat_protein
XX
PN   AU9211048-A.
XX
PD   02-SEP-1993.
XX
PF   18-FEB-1992;   92AU-0011048.
XX
PR   18-FEB-1992;   92AU-0011048.
XX
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Baeck AC,  Busch A,  Convents AC;
XX
DR   WPI; 1993-328419/42.
DR   N-PSDB; AAQ49941.
XX
PT   Detergent compsns., esp. for fabrics - contain surfactant,
PT   softening clay and high activity cellulase, partic. from Humicola
PT   insolens
XX
PS   Claim 5; Page 54-55; 71pp; English.
XX
CC   A new detergent comprises a high activity cellulase in combination
CC   with a softening clay.  The cellulase may be an endoglucanase
CC   enzyme derived from Humicola insolens (AAQ49941) or Fusarium oxysporum
CC   (AAQ49942).  The combination provides a higher than additive softening
CC   performance and excellent colour rejuvenation and whiteness
CC   maintenance for fabrics.
XX
SQ   Sequence   305 AA;

  Query Match             69.9%;  Score 1159;  DB 14;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 5.2e-81;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || ||||||||||||||: | |||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| |||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :    ||||:  |  |:      |   ||||::|||||| |:|||||||:|:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 13
AAY67388
ID   AAY67388 standard; Protein; 305 AA.
XX
AC   AAY67388;
XX
DT   25-APR-2000  (first entry)
XX
DE   H. insolens endoglucanase enzyme protein sequence.
XX
KW   Endoglucanase; cellulase enzyme; detergent composition; anion surfactant;
KW   degradation protection; liquid detergent; long term stability.
XX
OS   Humicola insolens.
XX
PN   EP633311-A1.
XX
PD   11-JAN-1995.
XX
PF   28-JUN-1993;   93EP-0870122.
XX
PR   28-JUN-1993;   93EP-0870122.
XX
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Herbots IMAJ,  Jansen MP;
XX
DR   WPI; 1995-038508/06.
DR   N-PSDB; AAZ60178.
XX
PT   Liq. detergent compsns. - contg. hydrophobic amine(s) for cellulase
PT   stabilisation
XX
PS   Claim 11; Page 18-19; 28pp; English.
XX
CC   This sequence represents the Humicola insolens endoglucanase enzyme
CC   protein sequence. The enzyme is a cellulase enzyme used in the liquid
CC   detergent composition of the invention. The detergent comprises an anion
CC   surfactant, a cellulase enzyme and a cellulase stabilising amount of an
CC   amine of the formula R1R2R3N; where R1 and R2 are independently H or a
CC   C2-C9 alkyl chain, and R3 is a C2-C9 alkyl chain or cyclohexyl or
CC   cyclopentyl or cycloheptyl. The amine prevents the cellulase enzyme from
CC   becoming degraded. The composition is used as a liquid detergent, and has
CC   long term stability, it is applicable to the protection of an cellulase
CC   and can be used in the presence of proteases.
XX
SQ   Sequence   305 AA;

  Query Match             69.9%;  Score 1159;  DB 16;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 5.2e-81;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || ||||||||||||||: | |||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCGPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| |||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :    ||||:  |  |:      |   ||||::|||||| |:|||||||:|:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 14
AAW01502
ID   AAW01502 standard; protein; 305 AA.
XX
```

```
AC   AAW01502;
XX
DT   26-FEB-1997  (first entry)
XX
DE   43 kD endoglucanase, EG V.
XX
KW   Detergent composition; cellulase; retaining-type activity; catalytic
KW   activity; cellotriose; particulate soil removal; colour clarification;
KW   cleaning; cellulose-containing fabric; cellubiohydrolase; endoglucanase;
XX
OS   Humicola insolens, DSM 1800.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= signal_peptide
FT   Protein         22..305
FT                   /label= mature_protein
XX
PN   WO9502675-A1.
XX
PD   26-JAN-1995.
XX
PF   07-JUL-1994;   94WO-DK00280.
XX
PR   11-OCT-1993;   93DK-0001135.
PR   12-JUL-1993;   93EP-0870131.
XX
PA   (NOVO ) NOVO-NORDISK AS.
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Convents AC,  Jeffreys B,  Schuelein M,  Tikhomirov DF;
XX
DR   WPI; 1995-067325/09.
XX
PT   Detergent compsn. contg. two cellulase components - the first
PT   removing soil particles and the second capable of colour
PT   clarification, useful in laundry compsns.
XX
PS   Claim 25; Page 66-67; 83pp; English.
XX
CC   Detergent compositions comprising: 1) a first cellulase component having
CC   retaining-type activity, pref. having a catalytic activity on cellotriose
CC   at pH 8.5 corresponding to kcat of at least 0.01 s<-1>, and capable of
CC   particulate soil removal; and 2) a second cellulase component having
CC   multiple domains comprising at least one non-catalytic domain attached to
CC   a catalytic domain, pref. having a catalytic activity on Red Avicel 7.5
CC   per 1 mg of cellulase protein higher than 10<-4> IU and being capable of
CC   colour clarification, where at least one of the cellulase components is a
CC   single (recombinant) component, are useful for cleaning and colour
CC   clarification of cellulose-containing fabrics. The second cellulase
CC   component can be an endoglucanase which is immunoreactive with an
CC   antibody raised against a highly purified ~43 kD endoglucanase
CC   derived from Humicola insolens, DSM 1800, and is esp. the present
CC   sequence, designated EG V.
XX
SQ   Sequence   305 AA;

  Query Match             69.9%;  Score 1159;  DB 16;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 5.2e-81;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || |||||||||||||: | |||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||:|||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
             ||  ||||| |||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :    ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 15
AAR88471
ID   AAR88471 standard; Protein; 305 AA.
XX
AC   AAR88471;
XX
DT   03-APR-1996  (first entry)
XX
DE   Alkaline endoglucanase Carezyme.
XX
KW   Alkaline endoglucanase; carezyme; cellulase; host cell;
KW   Fusarium graminearum; heterologous gene expression.
XX
OS   Humicola insolens DSM 1800.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= Sig_peptide
XX
PN   WO9600787-A1.
XX
PD   11-JAN-1996.
XX
PF   15-JUN-1995;   95WO-US07743.
XX
PR   15-MAR-1995;   95US-0404678.
PR   30-JUN-1994;   94US-0269449.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
XX
PI   Moyer DL,  Royer JC,  Shuster JR,  Yoder W;
XX
DR   WPI; 1996-077498/08.
DR   N-PSDB; AAT10182.
XX
PT   Non-toxic, non-toxigenic, non-pathogenic recombinant Fusarium host
PT   cell - used to produce heterologous proteins, pref. enzymes,
PT   hormones, growth factors or receptors
XX
PS   Claim 13; Page 22-24; 38pp; English.
XX
CC   The Humicola insolens DSM 1800 alkaline endoglucanase Carezyme
CC   (AAR88471), a derivative of a 43 kDa cellulase, is expressed in
CC   non-toxic, non-toxigenic, non-pathogenic, protease-deficient
CC   recombinant Fusarium graminearum ATCC 20334 carrying plasmid
CC   pDM151.  This plasmid includes the Carezyme gene (AAT10182)
CC   operably linked to promoter and terminator (AAT10184-85) sequences
CC   of the Fusarium oxysporum trypsin-like protease SP387 gene.
CC   Recombinant enzyme is produced at levels of 6.0 g/l host cell;
CC   over 90% of secreted protein is Carezyme.
XX
SQ   Sequence   305 AA;

  Query Match             69.9%;  Score 1159;  DB 17;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 5.2e-81;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58
```

```
Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || |||||||||||||||: | |||||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :||||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :     ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305
```

Search completed: February 24, 2003, 11:51:32
Job time : 86 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2003, 11:38:03 ; Search time 14 Seconds
(without alignments)
628.390 Million cell updates/sec

Title: US-10-007-521-12
Perfect score: 1659
Sequence: 1 MRSTPVLRTTLAAALPLVAS.....TTCVSGTTCQKLNDYYSQCL 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1659	100.0	299	2	US-08-872-437-2	Sequence 2, Appli
2	1659	100.0	299	3	US-08-651-136C-12	Sequence 12, Appl
3	1659	100.0	299	4	US-09-229-911A-12	Sequence 12, Appl
4	1178	71.0	305	4	US-09-230-222-1	Sequence 1, Appli
5	1159	69.9	305	1	US-08-090-013-2	Sequence 2, Appli
6	1159	69.9	305	1	US-08-081-328-2	Sequence 2, Appli
7	1159	69.9	305	1	US-08-232-249-2	Sequence 2, Appli
8	1159	69.9	305	2	US-08-921-426-8	Sequence 8, Appli
9	1159	69.9	305	2	US-08-833-642A-2	Sequence 2, Appli
10	1159	69.9	305	2	US-08-140-008A-4	Sequence 4, Appli
11	1159	69.9	305	2	US-08-836-340-1	Sequence 1, Appli
12	1159	69.9	305	2	US-08-389-423-2	Sequence 2, Appli
13	1159	69.9	305	3	US-08-816-915-8	Sequence 8, Appli
14	1159	69.9	305	4	US-09-189-060B-56	Sequence 56, Appl
15	1159	69.9	305	4	US-09-230-665-2	Sequence 2, Appli
16	1159	69.9	305	4	US-09-189-028-2	Sequence 2, Appli
17	1159	69.9	305	5	PCT-US95-07743-8	Sequence 8, Appli
18	1142.5	68.9	306	4	US-09-189-060B-66	Sequence 66, Appl
19	1127	67.9	286	4	US-09-254-733-3	Sequence 3, Appli
20	1125	67.8	285	4	US-09-230-225B-6	Sequence 6, Appli
21	1112	67.0	284	1	US-08-411-777-10	Sequence 10, Appl
22	1112	67.0	284	3	US-09-057-088-10	Sequence 10, Appl
23	1107.5	66.8	306	4	US-09-189-060B-68	Sequence 68, Appl
24	1077.5	64.9	308	4	US-09-189-060B-70	Sequence 70, Appl
25	1076.5	64.9	304	4	US-09-189-060B-72	Sequence 72, Appl
26	1045	63.0	307	4	US-09-189-060B-74	Sequence 74, Appl
27	1039	62.6	235	4	US-09-329-350-31	Sequence 31, Appl
28	1021	61.5	308	3	US-08-651-136C-6	Sequence 6, Appli
29	1021	61.5	308	4	US-09-229-911A-6	Sequence 6, Appli
30	994.5	59.9	297	3	US-08-651-136C-4	Sequence 4, Appli
31	994.5	59.9	297	4	US-09-229-911A-4	Sequence 4, Appli
32	968.5	58.4	234	4	US-09-230-665-6	Sequence 6, Appli
33	962	58.0	310	3	US-08-651-136C-22	Sequence 22, Appl
34	962	58.0	310	4	US-09-229-911A-22	Sequence 22, Appl
35	951	57.3	295	3	US-08-651-136C-8	Sequence 8, Appli
36	951	57.3	295	4	US-09-229-911A-8	Sequence 8, Appli
37	939.5	56.6	376	1	US-08-090-013-4	Sequence 4, Appli
38	939.5	56.6	376	1	US-08-081-328-4	Sequence 4, Appli
39	939.5	56.6	376	1	US-08-232-249-4	Sequence 4, Appli
40	939.5	56.6	376	2	US-08-833-642A-4	Sequence 4, Appli
41	939.5	56.6	376	2	US-08-389-423-4	Sequence 4, Appli
42	939.5	56.6	376	4	US-09-189-028-4	Sequence 4, Appli
43	934.5	56.3	376	4	US-09-230-665-4	Sequence 4, Appli
44	923.5	55.7	294	3	US-08-651-136C-24	Sequence 24, Appl
45	923.5	55.7	294	4	US-09-229-911A-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-872-437-2
; Sequence 2, Application US/08872437
; Patent No. 5958082
; GENERAL INFORMATION:
; APPLICANT: Lund, Henrik
; APPLICANT: Kalum, Lisbeth
; TITLE OF INVENTION: Garments With Considerable Variation In
; TITLE OF INVENTION: Abrasion Level
; FILE REFERENCE: 4888.200-US
; CURRENT APPLICATION NUMBER: US/08/872,437
; CURRENT FILING DATE: 1997-06-10
; EARLIER APPLICATION NUMBER: 1276/96
; EARLIER FILING DATE: 1996-11-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Thielavia terrestris
US-08-872-437-2

Query Match 100.0%; Score 1659; DB 2; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.6e-127;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60

Qy      61 LSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGP 120

Qy     121 VAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCDS 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 VAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCDS 180

Qy     181 FPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 FPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNG 240

Qy     241 GTGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
```

RESULT 2
US-08-651-136C-12
; Sequence 12, Application US/08651136C

```
; Patent No. 6001639
;  GENERAL INFORMATION:
;    APPLICANT:  Schulein, Martin
;    APPLICANT:  Andersen, Lene N.
;    APPLICANT:  Lassen, Soren F.
;    APPLICANT:  Kauppinen, Markus S.
;    APPLICANT:  Lange, Lene
;    APPLICANT:  Nielsen, Ruby I.
;    APPLICANT:  Ihara, Michiko
;    APPLICANT:  Takagi, Shinobu
;    TITLE OF INVENTION:  No. 6001639el Endoglucanases
;    NUMBER OF SEQUENCES:  109
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6001639o No. 6001639disk of No. 6001639th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,136C
;      FILING DATE:  21-MAY-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4366.200-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  299 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-651-136C-12

  Query Match             100.0%;  Score 1659;  DB 3;  Length 299;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-127;
  Matches  299;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60

Qy      61 LSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGP 120

Qy     121 VAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCDS 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 VAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCDS 180

Qy     181 FPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 FPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNG 240

Qy     241 GTGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299


RESULT 3
US-09-229-911A-12
; Sequence 12, Application US/09229911A
; Patent No. 6387690
;   GENERAL INFORMATION:
;        APPLICANT: Schulein, Martin
;                   Andersen, Lene N.
;                   Lassen, Soren F.
;                   Kauppinen, Markus S.
;                   Lange, Lene
;                   Nielsen, Ruby I.
;                   Ihara, Michiko
;                   Takagi, Shinobu
;        TITLE OF INVENTION: No. 6387690el Endoglucanases
;        NUMBER OF SEQUENCES: 109
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: No. 6387690o No. 6387690disk of No. 6387690th America, Inc.
;             STREET: 405 Lexington Avenue, 64th Floor
;             CITY: New York
;             STATE: New York
;             COUNTRY: United States of America
;             ZIP: 10174-6401
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/229,911A
;             FILING DATE: 13-Jan-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/651,136
;             FILING DATE: 21-MAY-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Lambiris, Elias J.
;             REGISTRATION NUMBER: 33,728
;             REFERENCE/DOCKET NUMBER: 4366.200-US
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 212-867-0123
;             TELEFAX: 212-878-9655
;   INFORMATION FOR SEQ ID NO: 12:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 299 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-229-911A-12

  Query Match             100.0%;  Score 1659;  DB 4;  Length 299;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-127;
  Matches  299;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60

Qy      61 LSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGP 120

Qy     121 VAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCDS 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 VAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCDS 180

Qy     181 FPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 FPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNG 240

Qy     241 GTGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299


RESULT 4
US-09-230-222-1
; Sequence 1, Application US/09230222A
```

```
; Patent No. 6159720
; GENERAL INFORMATION:
;  APPLICANT: MURASHIMA, KOUICHIROU
;  APPLICANT: MORIYA, TATSUKI
;  APPLICANT: HAMAYA, TORU
;  APPLICANT: KOGA, JINICHIRO
;  APPLICANT: SUMIDA, NAOMI
;  APPLICANT: AOYAGI, KAORU
;  APPLICANT: MURAKAMI, TAKESHI
;  APPLICANT: KONO, TOSHIAKI
;  TITLE OF INVENTION: ENZYME ENDOGLUCANASE AND CELLULASE PREPARATIONS
;  TITLE OF INVENTION: CONTAINING THE SAME
;  FILE REFERENCE: 99-0055*/LC(WMC)/144
;  CURRENT APPLICATION NUMBER: US/09/230,222A
;  CURRENT FILING DATE: 1999-03-03
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 305
;   TYPE: PRT
;   ORGANISM: Humicola insolens
US-09-230-222-1

  Query Match             71.0%;  Score 1178;  DB 4;  Length 305;
  Best Local Similarity   68.1%;  Pred. No. 2e-88;
  Matches  209;  Conservative   44;  Mismatches   44;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:||: : ||||::| ||   |:|||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLRSAVVAALPVLALAA--DGKSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           |:||: :|||  || ||||||||||||||: |:|||||||| :|: |||||| ||||||
Db      59 LTDFDAKSGCEPGGVAYSCADQTPWAVNDDFAFGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||:||: ||||||| :||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| |||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :   | ||||:  |  |:      |   |||:::||||||  |:|||||||:|:|| |:|
Db     239 SSPVGQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 5
US-08-090-013-2
; Sequence 2, Application US/08090013
; Patent No. 5443750
;  GENERAL INFORMATION:
;    APPLICANT:  CONVENTS, ANDRE C
;    APPLICANT:  BUSCH, ALFRED
;    APPLICANT:  BAECK, ANDRE C
;    TITLE OF INVENTION:  DETERGENT COMPOSITION WITH HIGH ACTIVITY
;    TITLE OF INVENTION:  CELLULASE AND SOFTENING CLAYS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THE PROCTER & GAMBLE COMPANY
;      STREET:  5299 SPRING GROVE AVENUE
;      CITY:  CINCINNATI
;      STATE:  OHIO
;      COUNTRY:  USA
;      ZIP:  45217
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/090,013
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 91202880.0
;      FILING DATE:  06-NOV-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PATEL, KEN K.
;      REGISTRATION NUMBER:  33,988
;      REFERENCE/DOCKET NUMBER:  CM393
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  513-627-6437
;      TELEFAX:  513-627-4854
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  305 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-090-013-2

  Query Match             69.9%;  Score 1159;  DB 1;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 7.1e-87;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:|||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || ||||||||||||||: | |||||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||:||: ||||||| :||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| |||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :     ||||:  |  |:      |   |||:::||||||  |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 6
US-08-081-328-2
; Sequence 2, Application US/08081328
; Patent No. 5520838
;  GENERAL INFORMATION:
;    APPLICANT:  BAECK, ANDRE C.
;    APPLICANT:  CEULEMANS, RAPHAEL ANGELINE A.
;    APPLICANT:  BUSCH, ALFRED (NMN)
;    TITLE OF INVENTION:  COMPACT DETERGENT COMPOSITIONS WITH
;    TITLE OF INVENTION:  HIGH ACTIVITY CELLULASE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THE PROCTER & GAMBLE COMPANY
;      STREET:  11810 East Miami River Road
;      CITY:  CINCINNATI
;      STATE:  OHIO
;      COUNTRY:  USA
;      ZIP:  45253-8707
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/081,328
;      FILING DATE:  11/19/93
;      CLASSIFICATION:  252
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ZERBY, KIM WILLIAM
;      REGISTRATION NUMBER:  32,323
;      REFERENCE/DOCKET NUMBER:  CM356M
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  513-627-2885
;      TELEFAX:  513-627-0318
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  305 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-081-328-2

  Query Match             69.9%;  Score 1159;  DB 1;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 7.1e-87;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:|||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || ||||||||||||||: | ||||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :     ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 7
US-08-232-249-2
; Sequence 2, Application US/08232249
; Patent No. 5610129
;  GENERAL INFORMATION:
;    APPLICANT:  McCORQUODALE, FINLAY (NMN)
;    APPLICANT:  BUSCH, ALFRED (NMN)
;    TITLE OF INVENTION:  DYE TRANSFER INHIBITING COMPOSITIONS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THE PROCTER & GAMBLE COMPANY
;      STREET:  5299 SPRING GROVE AVENUE
;      CITY:  CINCINNATI
;      STATE:  OHIO
;      COUNTRY:  USA
;      ZIP:  45217
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,249
;      FILING DATE:  05-MAY-1994
;      CLASSIFICATION:  510
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 91202882.6
;      FILING DATE:  06-NOV-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ALLEN, GEORGE W.
;      REGISTRATION NUMBER:  26,143
;      REFERENCE/DOCKET NUMBER:  CM-395
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  513-627-5946
;      TELEFAX:  513-627-8118
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  305 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-232-249-2

  Query Match             69.9%;  Score 1159;  DB 1;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 7.1e-87;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:|||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || ||||||||||||||: | ||||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :     ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 8
US-08-921-426-8
; Sequence 8, Application US/08921426
; Patent No. 5837847
;  GENERAL INFORMATION:
;    APPLICANT:  Royer, John C
;    APPLICANT:  Moyer, Donna L
;    APPLICANT:  Yoder, Wendy T
;    APPLICANT:  Shuster, Jeffrey R
;    TITLE OF INVENTION:  NON-TOXIC, NON-TOXIGENIC, NON-PATHOGENIC
;    TITLE OF INVENTION:  FUSARIUM EXPRESSION SYSTEM
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5837847o No. 5837847disk of No. 5837847th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/921,426
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/456,433
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  US 08/404,678
;      FILING DATE:  15-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris Dr., Cheryl H.
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  4216.010-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  305 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-921-426-8

  Query Match             69.9%;  Score 1159;  DB 2;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 7.1e-87;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy      1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
          |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db      1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy     61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
          ::||: :|||  || |||||||||||||: | |||||||| :|: |||||| ||||||
Db     59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy    120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
          ||||| |||||||||||||| ||: :||||||||:||: ||||||| :|||||||::||
Db    119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy    180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db    179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy    239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
          :     ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db    239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy    293 DYYSQCL 299
          |:| |||
Db    299 DWYHQCL 305


RESULT 9
US-08-833-642A-2
; Sequence 2, Application US/08833642A
; Patent No. 5883066
;  GENERAL INFORMATION:
;    APPLICANT:  Ivan M. A. J. Herbots et al.
;    TITLE OF INVENTION:  Liquid Detergent Compositions
;    TITLE OF INVENTION:  Containing Cellulase and Amine
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jackie Ann Zurcher
;      ADDRESSEE:  Dinsmore & Shohl LLP
;      STREET:  255 E. Fifth Street
;      STREET:  1900 Chemed Center
;      CITY:  Cincinnati
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  45202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/833,642A
;      FILING DATE:  April 8, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Zurcher, J. A.
;      REGISTRATION NUMBER:  P42,251
;      REFERENCE/DOCKET NUMBER:  CM551C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (513) 977-8377
;      TELEFAX:  (513) 977-8141
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  305 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-833-642A-2

  Query Match             69.9%;  Score 1159;  DB 2;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 7.1e-87;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy      1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
          |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db      1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy     61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
          ::||: :|||  || |||||||||||||: | |||||||| :|: |||||| ||||||
Db     59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy    120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
          ||||| |||||||||||||| ||: :||||||||:||: ||||||| :|||||||::||
Db    119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy    180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db    179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy    239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
          :     ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db    239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy    293 DYYSQCL 299
          |:| |||
Db    299 DWYHQCL 305


RESULT 10
US-08-140-008A-4
; Sequence 4, Application US/08140008A
; Patent No. 5914306
;  GENERAL INFORMATION:
;    APPLICANT:  SVENDSEN, Allan
;    APPLICANT:  VON DER OSTEN, Claus
;    APPLICANT:  CLAUSEN, Ib Groth
;    APPLICANT:  PATKAR, Shamkant Anant
;    APPLICANT:  BORCH, Kim
;    TITLE OF INVENTION:  STABILIZED ENZYMES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5914306o No. 5914306disk of No. 5914306th America
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/140,008A
;      FILING DATE:  22-OCT-1993
;      CLASSIFICATION:  252
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  3601.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 867-0123
;      TELEFAX:  (212) 878-9655
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  305 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-140-008A-4

  Query Match             69.9%;  Score 1159;  DB 2;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 7.1e-87;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || |||||||||||||: | |||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :||||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :     ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 11
US-08-836-340-1
; Sequence 1, Application US/08836340
; Patent No. 5916798
;  GENERAL INFORMATION:
;    APPLICANT:  Lund, Henrik
;    APPLICANT:  Pedersen, Hanne H st
;    TITLE OF INVENTION:  A Method of Obtaining a Cellulosic Textile
;    TITLE OF INVENTION:  Fabric with Reduced Tendency to Pilling Formation
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5916798o No. 5916798disk of No. 5916798th America
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,340
;      FILING DATE:
;      CLASSIFICATION:  435
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  305 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Humicola insolens
;      STRAIN:  DSM 1800
US-08-836-340-1

  Query Match             69.9%;  Score 1159;  DB 2;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 7.1e-87;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDDFALGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || |||||||||||||: | |||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :||||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :     ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 12
US-08-389-423-2
; Sequence 2, Application US/08389423
; Patent No. 5948672
;  GENERAL INFORMATION:
;    APPLICANT:  Rasmussen, Grethe
;    APPLICANT:  Mikkelsen, Jan Moller
;    APPLICANT:  Schulein, Martin
;    APPLICANT:  Patkar, Shankant A.
;    APPLICANT:  Hagen, Fred
;    TITLE OF INVENTION:  A Cellulase Preparation Comprising an
;    TITLE OF INVENTION:  Endoglucanase Enzyme
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5948672o No. 5948672disk of No. 5948672th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/389,423
;      FILING DATE:  14-FEB-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  3469.214-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  305 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-389-423-2

  Query Match            69.9%;  Score 1159;  DB 2;  Length 305;
  Best Local Similarity  67.1%;  Pred. No. 7.1e-87;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || |||||||||||||: | |||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :    ||||:  |  |:      |  ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 13
US-08-816-915-8
; Sequence 8, Application US/08816915
; Patent No. 6060305
;  GENERAL INFORMATION:
;    APPLICANT:  Royer, John C
;    APPLICANT:  Moyer, Donna L
;    APPLICANT:  Yoder, Wendy T
;    APPLICANT:  Shuster, Jeffrey R
;    TITLE OF INVENTION:  NON-TOXIC, NON-TOXIGENIC, NON-PATHOGENIC
;    TITLE OF INVENTION:  FUSARIUM EXPRESSION SYSTEM
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6060305o No. 6060305disk of No. 6060305th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/816,915
;      FILING DATE:  13-MAR-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris Dr., Cheryl H.
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  4216.240-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  305 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-816-915-8

  Query Match            69.9%;  Score 1159;  DB 3;  Length 305;
  Best Local Similarity  67.1%;  Pred. No. 7.1e-87;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :|||  || |||||||||||||: | |||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :    ||||:  |  |:      |  ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 14
US-09-189-060B-56
; Sequence 56, Application US/09189060B
; Patent No. 6270968
; GENERAL INFORMATION:
;  APPLICANT: Dalboge, Henrik
;  APPLICANT:  Sandal, Thomas
;  APPLICANT:  Kauppinen, Markus
;  APPLICANT:  Borge, Diderichsen
;  TITLE OF INVENTION: Method Of Providing No. 6270968el DNA Sequences
;  FILE REFERENCE: 4772.204-US
;  CURRENT APPLICATION NUMBER: US/09/189,060B
;  CURRENT FILING DATE:  1998-11-10
;  PRIOR APPLICATION NUMBER: PCT/DK97/00216
;  PRIOR FILING DATE: 1997-05-12
;  NUMBER OF SEQ ID NOS: 74
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 56
;   LENGTH: 305
;   TYPE: PRT
;   ORGANISM: Humicola insolens
;   FEATURE:
;   NAME/KEY: SIGNAL
;   LOCATION: (1)...(21)
US-09-189-060B-56

  Query Match            69.9%;  Score 1159;  DB 4;  Length 305;
  Best Local Similarity  67.1%;  Pred. No. 7.1e-87;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
```

```
          |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:||||||
Db      1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy     61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
          ::||: :|||  || |||||||||||||: | |||||||||| :|: |||||| ||||||
Db     59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy    120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
          ||||| |||||||||||||| ||: :|||||||:||: ||||||| :|||||||::||
Db    119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy    180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
           ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db    179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy    239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
          :    ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db    239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy    293 DYYSQCL 299
          |:| |||
Db    299 DWYHQCL 305


RESULT 15
US-09-230-665-2
; Sequence 2, Application US/09230665
; Patent No. 6322595
; GENERAL INFORMATION:
;  APPLICANT: Boyer, Stanton L
;  TITLE OF INVENTION: Detergent Composition Comprising Two Cellulase
;  TITLE OF INVENTION: Components, with and without a Cellulose-Binding Domain
;  FILE REFERENCE: 6191 Sequence Listing (8 Sequences)
; Patent No. 6322595
;  CURRENT APPLICATION NUMBER: US/09/230,665
;  CURRENT FILING DATE: 1999-01-29
;  EARLIER APPLICATION NUMBER: 60/023,125
;  EARLIER FILING DATE: 1996-07-30
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 305
;   TYPE: PRT
;   ORGANISM: Humicola insolens
US-09-230-665-2

  Query Match             69.9%;  Score 1159;  DB 4;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 7.1e-87;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy      1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
          |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:||||||
Db      1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy     61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
          ::||: :|||  || |||||||||||||: | |||||||||| :|: |||||| ||||||
Db     59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy    120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
          ||||| |||||||||||||| ||: :|||||||:||: ||||||| :|||||||::||
Db    119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy    180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
           ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db    179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy    239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
          :    ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db    239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy    293 DYYSQCL 299
          |:| |||
Db    299 DWYHQCL 305
```

Search completed: February 24, 2003, 11:43:05
Job time : 15 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 24, 2003, 11:42:08 ; Search time 13 Seconds
(without alignments)
714.606 Million cell updates/sec

Title: US-10-007-521-12
Perfect score: 1659
Sequence: 1 MRSTPVLRTTLAAALPLVAS.....TTCVSGTTCQKLNDYYSQCL 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 156504 seqs, 31069816 residues

Total number of hits satisfying chosen parameters: 156504

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1159	69.9	305	10	US-09-735-787-2	Sequence 2, Appli
2	1039	62.6	235	1	US-08-841-636A-31	Sequence 31, Appl
3	939.5	56.6	376	10	US-09-735-787-4	Sequence 4, Appli
4	160.5	9.7	480	10	US-09-823-936-16	Sequence 16, Appl
5	160	9.6	33	9	US-09-916-494A-6	Sequence 6, Appli
6	158	9.5	34	1	US-08-841-636A-37	Sequence 37, Appl
7	154	9.3	4440	9	US-10-174-590-525	Sequence 525, App
8	154	9.3	4440	9	US-10-176-758-525	Sequence 525, App
9	154	9.3	4440	9	US-10-175-737-525	Sequence 525, App
10	154	9.3	4440	9	US-10-173-706-525	Sequence 525, App
11	154	9.3	4440	9	US-10-175-738-525	Sequence 525, App
12	154	9.3	4440	9	US-10-175-752-525	Sequence 525, App
13	154	9.3	4440	9	US-10-176-482-525	Sequence 525, App
14	154	9.3	4440	9	US-10-176-757-525	Sequence 525, App
15	154	9.3	4440	9	US-10-176-913-525	Sequence 525, App
16	154	9.3	4440	9	US-10-180-552-525	Sequence 525, App
17	154	9.3	4440	9	US-10-180-557-525	Sequence 525, App
18	154	9.3	4440	9	US-10-173-700-525	Sequence 525, App
19	154	9.3	4440	9	US-10-174-572-525	Sequence 525, App
20	154	9.3	4440	9	US-10-174-579-525	Sequence 525, App
21	154	9.3	4440	9	US-10-174-582-525	Sequence 525, App
22	154	9.3	4440	9	US-10-174-588-525	Sequence 525, App
23	154	9.3	4440	9	US-10-175-739-525	Sequence 525, App
24	154	9.3	4440	9	US-10-175-740-525	Sequence 525, App
25	154	9.3	4440	9	US-10-175-743-525	Sequence 525, App
26	154	9.3	4440	9	US-10-176-488-525	Sequence 525, App
27	154	9.3	4440	9	US-10-176-492-525	Sequence 525, App
28	154	9.3	4440	9	US-10-176-747-525	Sequence 525, App
29	154	9.3	4440	9	US-10-176-750-525	Sequence 525, App
30	154	9.3	4440	9	US-10-176-985-525	Sequence 525, App
31	154	9.3	4440	9	US-10-176-987-525	Sequence 525, App
32	154	9.3	4440	9	US-10-176-991-525	Sequence 525, App
33	154	9.3	4440	9	US-10-176-992-525	Sequence 525, App
34	154	9.3	4440	9	US-10-176-993-525	Sequence 525, App
35	154	9.3	4440	9	US-10-184-658-525	Sequence 525, App
36	154	9.3	4440	9	US-10-173-695-525	Sequence 525, App
37	154	9.3	4440	9	US-10-173-697-525	Sequence 525, App
38	154	9.3	4440	9	US-10-173-705-525	Sequence 525, App
39	154	9.3	4440	9	US-10-174-576-525	Sequence 525, App
40	154	9.3	4440	9	US-10-174-585-525	Sequence 525, App
41	154	9.3	4440	9	US-10-174-586-525	Sequence 525, App
42	154	9.3	4440	9	US-10-175-747-525	Sequence 525, App
43	154	9.3	4440	9	US-10-176-481-525	Sequence 525, App
44	154	9.3	4440	9	US-10-176-485-525	Sequence 525, App
45	154	9.3	4440	9	US-10-176-487-525	Sequence 525, App

ALIGNMENTS

RESULT 1
US-09-735-787-2
; Sequence 2, Application US/09735787
; Patent No. US20010036910A1
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Grethe
; Mikkelsen, Jan Moller
; Schulein, Martin
; Patkar, Shankant A.
; Hagen, Fred
; TITLE OF INVENTION: A Cellulase Preparation Comprising an
; Endoglucanase Enzyme
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. US20010036910A1o No. US20010036910A1disk of No. US200100
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/735,787
; FILING DATE: 13-Dec-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/189,028
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3469.214-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 305 amino acids

```
;             TYPE: amino acid
;             TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-735-787-2

  Query Match             69.9%;  Score 1159;  DB 10;  Length 305;
  Best Local Similarity   67.1%;  Pred. No. 3.7e-83;
  Matches  206;  Conservative   44;  Mismatches   47;  Indels   10;  Gaps    4;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           |||:|:| : : ||||::| ||   |:||||||||||||| |  || |:|||::|:|||||
Db       1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58

Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           ::||: :||| || ||||||||||||: | |||||||| :|: |||||| ||||||
Db      59 ITDFDAKSGCEPGGVAYSCADQTPWAVNDDFALGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :||||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| ||||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :    ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVNQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 2
US-08-841-636A-31
; Sequence 31, Application US/08841636A
; Patent No. US20020168751A1
;  GENERAL INFORMATION:
;    APPLICANT:  Miettinen-Oinonen, Arja
;    APPLICANT:  Londesborough, John
;    APPLICANT:  Vehmaanper , Jari
;    APPLICANT:  Haakana, Heli
;    APPLICANT:  M ntyl , Arja
;    APPLICANT:  Lantto, Raija
;    APPLICANT:  Elovainio, Minna
;    APPLICANT:  Joutsjoki, Vesa
;    APPLICANT:  Paloheimo, Marja
;    APPLICANT:  Suominen, Pirkko
;    TITLE OF INVENTION:  NOVEL CELLULASES, THE GENES ENCODING THEM AND
;    TITLE OF INVENTION:  USES THEREOF
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/841,636A
;      FILING DATE:  30-APR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/005,335
;      FILING DATE:  17-OCT-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/007,926
;      FILING DATE:  04-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/020,840
;      FILING DATE:  28-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/732,181
;      FILING DATE:  16-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FI96/00550
;      FILING DATE:  17-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Timothy J. Shea, Jr.
;      REGISTRATION NUMBER:  41,306
;      REFERENCE/DOCKET NUMBER:  1716.0510005/MAC/TJS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)371-2600
;      TELEFAX:  (202)371-2540
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  235 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Melanocarpus albomyces
;      STRAIN:  ALKO4237
;    FEATURE:
;      NAME/KEY:  Protein
;      LOCATION:  1..235
;      OTHER INFORMATION:  /label= 20K-cellulase
US-08-841-636A-31

  Query Match             62.6%;  Score 1039;  DB 1;  Length 235;
  Best Local Similarity   76.3%;  Pred. No. 6.1e-74;
  Matches  180;  Conservative   27;  Mismatches   27;  Indels    2;  Gaps    1;

Qy       1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
           ||||||||  |||||||  | ||  :|||||||||||||| | ||  |:||||:|||||||
Db       1 MRSTPVLRALLAAALPLGALAA--NGQSTRYWDCCKPSCGWRGKGPVNQPVYSCDANFQR 58

Qy      61 LSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGP 120
           : ||:  ||| || |:|||| :|||:||||:|||||:::| :| ||||||||||||||
Db      59 IHDFDAVSGCEGGPAFSCADHSPWAINDNLSYGFAATALSGQTEESWCCACYALTFTSGP 118

Qy     121 VAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCDS 180
           |||||||||||||||||| ||: :||||||:|:||: |||||||:||||||| :|||
Db     119 VAGKTMVVQSTSTGGDLGSNHFDLNIPGGGVGLFDGCTPQFGGLPGARYGGISSRQECDS 178

Qy     181 FPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPS 236
           || |||||||||||||||||:|||::|||| |:|||:||:|:||  | ||  ||
Db     179 FPEPLKPGCQWRFDWFQNADNPSFTFERVQCPEELVARTGCRRHDDGGFAVFKAPS 234


RESULT 3
US-09-735-787-4
; Sequence 4, Application US/09735787
; Patent No. US20010036910A1
;   GENERAL INFORMATION:
;        APPLICANT: Rasmussen, Grethe
;                   Mikkelsen, Jan Moller
;                   Schulein, Martin
;                   Patkar, Shankant A.
;                   Hagen, Fred
;        TITLE OF INVENTION: A Cellulase Preparation Comprising an
;                            Endoglucanase Enzyme
;        NUMBER OF SEQUENCES: 33
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: No. US20010036910A1o No. US20010036910A1disk of No. US200100
;             STREET: 405 Lexington Avenue, 64th Floor
;             CITY: New York
;             STATE: New York
```

```
;             COUNTRY: United States of America
;             ZIP: 10174-6401
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/735,787
;             FILING DATE: 13-Dec-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/189,028
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Lambiris, Elias J.
;             REGISTRATION NUMBER: 33,728
;             REFERENCE/DOCKET NUMBER: 3469.214-US
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 212-867-0123
;             TELEFAX: 212-878-9655
;    INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 376 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-735-787-4

  Query Match             56.6%;  Score 939.5;  DB 10;  Length 376;
  Best Local Similarity   49.5%;  Pred. No. 5.5e-66;
  Matches  183;  Conservative   35;  Mismatches   73;  Indels   79;  Gaps    6;

Qy       9 TTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSDFNVQS 68
           | || | ||  |||||||  |||||||||||||:| |||||: |   || |   :|: |  :
Db       5 TLLALAGPLAVSAASGSGHSTRYWDCCKPSCSWSGKAAVNAPALTCDKNDNPISNTNAVN 64

Qy      69 GC-NGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMV 127
             ||  |||||:| : :|||||| ||||||||  |:||||:|||||||||||||:||| || |:
Db      65 GCEGGGSAYACTNYSPWAVNDELAYGFAATKISGGSEASWCCACYALTFTTGPVKGKKMI 124

Qy     128 VQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFG-GLPGAQYGGISSRDQCDSFPAPLK 186
           ||||:|||||| | ||: |||||||||:|||:|:|| | |||||||||| :|||:|  ||
Db     125 VQSTNTGGDLGDNHFDLMMPGGGVGIFDGCTSEFGKALGGAQYGGISSRSECDSYPELLK 184

Qy     187 PGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVF-------------- 232
            || ||||||:||||| |||:|||||  ::  |||||:|||||| |
Db     185 DGCHWRFDWFENADNPDFTFEQVQCPKALLDISGCKRDDDSSFPAFKVDTSASKPQPSSS 244

Qy     233 ---------------------------TPPS--------GGNGGTGTPTSTAPGSGQTSP 257
                                       | |:          |  | :| | : :
Db     245 AKKTTSAAAAAQPQKTKDSAPVVQKSSTKPAAQPEPTKPADKPQTDKPVATKPAATKPVQ 304

Qy     258 GGGSGCTSQK-------------------------WAQCGGIGF---SGCTTCVSGTTCQ 289
                 |:||                         : ||||      :|   | :|: |
Db     305 PVNKPKTTQKVRGTKTRGSCPAKTDATAKASVVPAYYQCGGSKSAYPNGNLACATGSKCV 364

Qy     290 KLNDYYSQCL 299
           | |:||||||:
Db     365 KQNEYYSQCV 374


RESULT 4
US-09-823-936-16
; Sequence 16, Application US/09823936
; Patent No. US20020061309A1
; GENERAL INFORMATION:
;  APPLICANT: GARGER, Stephen J.
;  APPLICANT:  GROSS,  Cynthia
;  APPLICANT:  LINDBO, John A.
;  APPLICANT:  POGUE, Gregory P.
;  TITLE OF INVENTION: PRODUCTION OF PEPTIDES IN PLANTS AS
;  TITLE OF INVENTION:  N-TERMINAL VIRAL COAT PROTEIN FUSIONS
;  FILE REFERENCE: 008010087CPUS05
;  CURRENT APPLICATION NUMBER: US/09/823,936
;  CURRENT FILING DATE:  2001-03-29
;  PRIOR APPLICATION NUMBER: 09/ 520,967
;  PRIOR FILING DATE: 2000-03-08
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
;   LENGTH: 480
;   TYPE: PRT
;   ORGANISM: P. Yoeli
US-09-823-936-16

  Query Match             9.7%;  Score 160.5;  DB 10;  Length 480;
  Best Local Similarity   23.9%;  Pred. No. 2.9e-05;
  Matches   71;  Conservative   16;  Mismatches  119;  Indels   91;  Gaps   11;

Qy      19 ASAASG-SGQSTRYWDCC----KPSCAWPGKAAVSQP------VYACDANFQRLSDFNVQ 67
             | ||:|  | :|    ||      :||  | |              | |
Db     218 ACAATGCGGTATTAGACCCGCTAGTCACAGCACTGTTAGGTGCATTCGA---------- 266

Qy      68 SGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGG-SESSWCCACYALTFTSGPVAGKTM 126
             |  |:| :    |           | || : |||   :  |||| | |   |  |  |
Db     267 CACTAGAAATAGAATAATAGAAGTTGAAAATCAGGCGAACCCCACGACTGCCGAAACGT- 325

Qy     127 VVQSTSTGGDLGSNQFDIAMPGGG-VGIFNGCS----SQFGGLPGAQYGGISSRDQCDSF 181
                | |  |          | |  |   |          ||      ||
Db     326 ------TAGATGCTACTCGTAGAGTAGACGACGCAACGGTGGCCATAAGG---------- 369

Qy     182 PAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNGG 241
                 ||                                ||     ::    |  : |  |
Db     370 -----AGC----------------------------GCGATAAATAATTTAATAGTAG 394

Qy     242 TGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFS----------GCTTCVSGTT 287
             |  :|  |:|  :  ||: ||:   | ||| | :           ||||  |||
Db     395 AATTGATCAGAGGAACCGGATCTTATAATCGGAGCTCTTTCGAGAGCTCTTCTGGTT 451


RESULT 5
US-09-916-494A-6
; Sequence 6, Application US/09916494A
; Patent No. US20020164774A1
; GENERAL INFORMATION:
;  APPLICANT: Fowler, Timothy
;  APPLICANT:  Clarkson, Kathleen A.
;  APPLICANT:  Ward, Michael
;  APPLICANT:  Collier, Katherine D.
;  APPLICANT:  Larenas, Edmund
;  TITLE OF INVENTION: Method and Compositions for Treating
;  TITLE OF INVENTION:  Cellulose Containing Fabrics Using Truncated Cellulase
;  TITLE OF INVENTION:  Enzyme Compositions
;  FILE REFERENCE: GC226-C4
;  CURRENT APPLICATION NUMBER: US/09/916,494A
;  CURRENT FILING DATE:  2000-06-14
;  PRIOR APPLICATION NUMBER: US 08/382,452
;  PRIOR FILING DATE: 1995-02-01
;  PRIOR APPLICATION NUMBER: US 08/169,948
;  PRIOR FILING DATE: 1993-12-17
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 33
;   TYPE: PRT
;   ORGANISM: Trichoderma longibrachiatum
US-09-916-494A-6

  Query Match             9.6%;  Score 160;  DB 9;  Length 33;
  Best Local Similarity   81.2%;  Pred. No. 1.8e-06;
  Matches   26;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy     268 WAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           | |||||||:||| || |||||||  ||||||||
```

```
Db      2 WGQCGGIGYSGCKTCTSGTTCQYSNDYYSQCL 33


RESULT 6
US-08-841-636A-37
; Sequence 37, Application US/08841636A
; Patent No. US20020168751A1
;  GENERAL INFORMATION:
;    APPLICANT:  Miettinen-Oinonen, Arja
;    APPLICANT:  Londesborough, John
;    APPLICANT:  Vehmaanper , Jari
;    APPLICANT:  Haakana, Heli
;    APPLICANT:  M ntyl , Arja
;    APPLICANT:  Lantto, Raija
;    APPLICANT:  Elovainio, Minna
;    APPLICANT:  Joutsjoki, Vesa
;    APPLICANT:  Paloheimo, Marja
;    APPLICANT:  Suominen, Pirkko
;    TITLE OF INVENTION:  NOVEL CELLULASES, THE GENES ENCODING THEM AND
;    TITLE OF INVENTION:  USES THEREOF
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/841,636A
;      FILING DATE:  30-APR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/005,335
;      FILING DATE:  17-OCT-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/007,926
;      FILING DATE:  04-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/020,840
;      FILING DATE:  28-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/732,181
;      FILING DATE:  16-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FI96/00550
;      FILING DATE:  17-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Timothy J. Shea, Jr.
;      REGISTRATION NUMBER:  41,306
;      REFERENCE/DOCKET NUMBER:  1716.0510005/MAC/TJS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)371-2600
;      TELEFAX:  (202)371-2540
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  34 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    ORIGINAL SOURCE:
;      ORGANISM:  Melanocarpus albomyces
;      STRAIN:  ALKO4237
;    FEATURE:
;      NAME/KEY:  Protein
;      LOCATION:  1..34
;      OTHER INFORMATION:  /label= prot-with-CBD
US-08-841-636A-37

  Query Match             9.5%;  Score 158;  DB 1;  Length 34;
  Best Local Similarity   73.5%;  Pred. No. 2.6e-06;
  Matches   25;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy    266 QKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
          ||| |||||||:|||||| :|:|:|    |:|||||
Db      1 QKWGQCGGIGYSGCTTCKAGSTCPAQNEYYSQCL 34


RESULT 7
US-10-174-590-525
; Sequence 525, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C42
;  CURRENT APPLICATION NUMBER: US/10/174,590
;  CURRENT FILING DATE:  2002-06-18
;  Prior application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
;   LENGTH: 4440
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-174-590-525

  Query Match             9.3%;  Score 154;  DB 9;  Length 4440;
  Best Local Similarity   24.8%;  Pred. No. 0.001;
  Matches   71;  Conservative   19;  Mismatches  114;  Indels   82;  Gaps   14;

Qy      9 TTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSDFNVQS 68
            |   |       |:|:|  |     |  ||           ||        :    :
Db    164 TGTGTATGCATGCATGTGTGTG----CGTGCA----------CACGTG----TGTGTTT 204

Qy     69 GCNGGSAYSCADQTPWAVNDNLAYGF--AATSIAGGSESSWCCACYALTFTSGPVAGKTM 126
          ||  |:|  ||  |  |         |     |    |: :   |||   | |:|  ||
Db    205 GCGTGTA--CATGTGCATGTGTGTGTTGTGTGCCTGTGTGTGCAC--ATGTAGAAAGAAA 260

Qy    127 VVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQ---FGGLPGAQYGGISSRDQCDSFPA 183
             :| ||   |:          || |    | :     ||  ||  |  :    |
Db    261 GTGATGTGTGGGA-------ATGGAGAATTCCAACCCAGGAGGA--GACTGTGCCTG--- 308

Qy    184 PLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNGGTG 243
            ||                                :||             : | ||||
Db    309 --GGGC-----------------------------TGC--------CACGAGAAGTGGTG 329

Qy    244 TPTSTAPGSGQTSPGGGSG-CTSQKWAQCGGIGFSGCTTCVSGTTC 288
            |  |:|  : |||:  | :     ||  |   |||| :||||
Db    330 ACACTTCGAGAAACGGGTTCCAAATGCACGTTGC--CTTCCAGTTC 373


RESULT 8
US-10-176-758-525
; Sequence 525, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
```

```
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C104
;  CURRENT APPLICATION NUMBER: US/10/176,758
;  CURRENT FILING DATE:  2002-06-21
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
;   LENGTH: 4440
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-176-758-525

  Query Match             9.3%;  Score 154;  DB 9;  Length 4440;
  Best Local Similarity   24.8%;  Pred. No. 0.001;
  Matches   71;  Conservative   19;  Mismatches  114;  Indels   82;  Gaps   14;

Qy       9 TTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSDFNVQS 68
             |    |       |:|:|  |     |   ||            ||       :    :
Db     164 TGTGTATGCATGCATGTGTGTG----CGTGCA----------CACGTG----TGTGTTT 204

Qy      69 GCNGGSAYSCADQTPWAVNDNLAYGF--AATSIAGGSESSWCCACYALTFTSGPVAGKTM 126
           ||  |:|  ||  |  |       |     |    |: :   |||   | |:|  ||
Db     205 GCGTGTA--CATGTGCATGTGTGTGTTGTGTGCCTGTGTGTGCAC--ATGTAGAAAGAAA 260

Qy     127 VVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQ---FGGLPGAQYGGISSRDQCDSFPA 183
           :| ||   |:         || |    | :     ||  ||  |  :    |
Db     261 GTGATGTGTGGGA-------ATGGAGAATTCCAACCCAGGAGGA--GACTGTGCCTG--- 308

Qy     184 PLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNGGTG 243
           ||                              :||             : | ||||
Db     309 --GGGC----------------------------TGC--------CACGAGAAGTGGTG 329

Qy     244 TPTSTAPGSGQTSPGGGSG-CTSQKWAQCGGIGFSGCTTCVSGTTC 288
           |  |:|  : |||:  | :     ||  |   |||| :||||
Db     330 ACACTTCGAGAAACGGGTTCCAAATGCACGTTGC--CTTCCAGTTC 373


RESULT 9
US-10-175-737-525
; Sequence 525, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C50
;  CURRENT APPLICATION NUMBER: US/10/175,737
;  CURRENT FILING DATE:  2002-06-19
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
;   LENGTH: 4440
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-175-737-525

  Query Match             9.3%;  Score 154;  DB 9;  Length 4440;
  Best Local Similarity   24.8%;  Pred. No. 0.001;
  Matches   71;  Conservative   19;  Mismatches  114;  Indels   82;  Gaps   14;

Qy       9 TTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSDFNVQS 68
             |    |       |:|:|  |     |   ||            ||       :    :
Db     164 TGTGTATGCATGCATGTGTGTG----CGTGCA----------CACGTG----TGTGTTT 204

Qy      69 GCNGGSAYSCADQTPWAVNDNLAYGF--AATSIAGGSESSWCCACYALTFTSGPVAGKTM 126
           ||  |:|  ||  |  |       |     |    |: :   |||   | |:|  ||
Db     205 GCGTGTA--CATGTGCATGTGTGTGTTGTGTGCCTGTGTGTGCAC--ATGTAGAAAGAAA 260

Qy     127 VVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQ---FGGLPGAQYGGISSRDQCDSFPA 183
           :| ||   |:         || |    | :     ||  ||  |  :    |
Db     261 GTGATGTGTGGGA-------ATGGAGAATTCCAACCCAGGAGGA--GACTGTGCCTG--- 308

Qy     184 PLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNGGTG 243
           ||                              :||             : | ||||
Db     309 --GGGC----------------------------TGC--------CACGAGAAGTGGTG 329

Qy     244 TPTSTAPGSGQTSPGGGSG-CTSQKWAQCGGIGFSGCTTCVSGTTC 288
           |  |:|  : |||:  | :     ||  |   |||| :||||
Db     330 ACACTTCGAGAAACGGGTTCCAAATGCACGTTGC--CTTCCAGTTC 373


RESULT 10
US-10-173-706-525
; Sequence 525, Application US/10173706
; Publication No. US20030022293A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C7
;  CURRENT APPLICATION NUMBER: US/10/173,706
;  CURRENT FILING DATE:  2002-06-17
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
;   LENGTH: 4440
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-173-706-525

  Query Match             9.3%;  Score 154;  DB 9;  Length 4440;
  Best Local Similarity   24.8%;  Pred. No. 0.001;
  Matches   71;  Conservative   19;  Mismatches  114;  Indels   82;  Gaps   14;

Qy       9 TTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSDFNVQS 68
             |    |       |:|:|  |     |   ||            ||       :    :
Db     164 TGTGTATGCATGCATGTGTGTG----CGTGCA----------CACGTG----TGTGTTT 204

Qy      69 GCNGGSAYSCADQTPWAVNDNLAYGF--AATSIAGGSESSWCCACYALTFTSGPVAGKTM 126
           ||  |:|  ||  |  |       |     |    |: :   |||   | |:|  ||
Db     205 GCGTGTA--CATGTGCATGTGTGTGTTGTGTGCCTGTGTGTGCAC--ATGTAGAAAGAAA 260

Qy     127 VVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQ---FGGLPGAQYGGISSRDQCDSFPA 183
           :| ||   |:         || |    | :     ||  ||  |  :    |
Db     261 GTGATGTGTGGGA-------ATGGAGAATTCCAACCCAGGAGGA--GACTGTGCCTG--- 308

Qy     184 PLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNGGTG 243
```

```
              ||                              :||             : | ||||
Db    309 --GGGC------------------------------TGC--------CACGAGAAGTGGTG 329

Qy    244 TPTSTAPGSGQTSPGGGSG-CTSQKWAQCGGIGFSGCTTCVSGTTC 288
              |  |:|  : |||:  | :     ||  |   |||| :||||
Db    330 ACACTTCGAGAAACGGGTTCCAAATGCACGTTGC--CTTCCAGTTC 373


RESULT 11
US-10-175-738-525
; Sequence 525, Application US/10175738
; Publication No. US20030022294A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C45
;  CURRENT APPLICATION NUMBER: US/10/175,738
;  CURRENT FILING DATE:  2002-06-19
;  Prior application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
;   LENGTH: 4440
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-175-738-525

  Query Match             9.3%;  Score 154;  DB 9;  Length 4440;
  Best Local Similarity   24.8%;  Pred. No. 0.001;
  Matches   71;  Conservative   19;  Mismatches  114;  Indels   82;  Gaps   14;

Qy      9 TTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSDFNVQS 68
            |    |       |:|:|  |     |   ||            ||       :     :
Db    164 TGTGTATGCATGCATGTGTGTG----CGTGCA-----------CACGTG----TGTGTTT 204

Qy     69 GCNGGSAYSCADQTPWAVNDNLAYGF--AATSIAGGSESSWCCACYALTFTSGPVAGKTM 126
           ||  |:|  ||  |  |       |     |    |: :   |||   | |:|  ||
Db    205 GCGTGTA--CATGTGCATGTGTGTGTTGTGTGCCTGTGTGTGCAC--ATGTAGAAAGAAA 260

Qy    127 VVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQ---FGGLPGAQYGGISSRDQCDSFPA 183
             :| ||   |:         || |    | :     ||  ||  |  :    |
Db    261 GTGATGTGTGGGA-------ATGGAGAATTCCAACCCAGGAGGA--GACTGTGCCTG--- 308

Qy    184 PLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNGGTG 243
              ||                              :||             : | ||||
Db    309 --GGGC------------------------------TGC--------CACGAGAAGTGGTG 329

Qy    244 TPTSTAPGSGQTSPGGGSG-CTSQKWAQCGGIGFSGCTTCVSGTTC 288
              |  |:|  : |||:  | :     ||  |   |||| :||||
Db    330 ACACTTCGAGAAACGGGTTCCAAATGCACGTTGC--CTTCCAGTTC 373


RESULT 12
US-10-175-752-525
; Sequence 525, Application US/10175752
; Publication No. US20030022295A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C60
;  CURRENT APPLICATION NUMBER: US/10/175,752
;  CURRENT FILING DATE:  2002-06-19
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
;   LENGTH: 4440
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-175-752-525

  Query Match             9.3%;  Score 154;  DB 9;  Length 4440;
  Best Local Similarity   24.8%;  Pred. No. 0.001;
  Matches   71;  Conservative   19;  Mismatches  114;  Indels   82;  Gaps   14;

Qy      9 TTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSDFNVQS 68
            |    |       |:|:|  |     |   ||            ||       :     :
Db    164 TGTGTATGCATGCATGTGTGTG----CGTGCA-----------CACGTG----TGTGTTT 204

Qy     69 GCNGGSAYSCADQTPWAVNDNLAYGF--AATSIAGGSESSWCCACYALTFTSGPVAGKTM 126
           ||  |:|  ||  |  |       |     |    |: :   |||   | |:|  ||
Db    205 GCGTGTA--CATGTGCATGTGTGTGTTGTGTGCCTGTGTGTGCAC--ATGTAGAAAGAAA 260

Qy    127 VVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQ---FGGLPGAQYGGISSRDQCDSFPA 183
             :| ||   |:         || |    | :     ||  ||  |  :    |
Db    261 GTGATGTGTGGGA-------ATGGAGAATTCCAACCCAGGAGGA--GACTGTGCCTG--- 308

Qy    184 PLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNGGTG 243
              ||                              :||             : | ||||
Db    309 --GGGC------------------------------TGC--------CACGAGAAGTGGTG 329

Qy    244 TPTSTAPGSGQTSPGGGSG-CTSQKWAQCGGIGFSGCTTCVSGTTC 288
              |  |:|  : |||:  | :     ||  |   |||| :||||
Db    330 ACACTTCGAGAAACGGGTTCCAAATGCACGTTGC--CTTCCAGTTC 373


RESULT 13
US-10-176-482-525
; Sequence 525, Application US/10176482
; Publication No. US20030022296A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C70
;  CURRENT APPLICATION NUMBER: US/10/176,482
;  CURRENT FILING DATE:  2002-06-20
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
;   LENGTH: 4440
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-176-482-525

  Query Match             9.3%;  Score 154;  DB 9;  Length 4440;
```

```
  Best Local Similarity   24.8%;  Pred. No. 0.001;
  Matches   71;  Conservative   19;  Mismatches  114;  Indels   82;  Gaps   14;

Qy       9 TTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSDFNVQS 68
           |    |       |:|:|  |     |   ||           ||       :    :
Db     164 TGTGTATGCATGCATGTGTGTG----CGTGCA----------CACGTG----TGTGTTT 204

Qy      69 GCNGGSAYSCADQTPWAVNDNLAYGF--AATSIAGGSESSWCCACYALTFTSGPVAGKTM 126
           ||  |:|  ||  |  |       |     |    |: :   |||   | |:|  ||
Db     205 GCGTGTA--CATGTGCATGTGTGTGTTGTGTGCCTGTGTGTGCAC--ATGTAGAAAGAAA 260

Qy     127 VVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQ---FGGLPGAQYGGISSRDQCDSFPA 183
             :| ||   |:         || |    | :     ||  ||  |  :    |
Db     261 GTGATGTGTGGGA-------ATGGAGAATTCCAACCCAGGAGGA--GACTGTGCCTG--- 308

Qy     184 PLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNGGTG 243
           ||                                :||                : | ||||
Db     309 --GGGC----------------------------TGC--------CACGAGAAGTGGTG 329

Qy     244 TPTSTAPGSGQTSPGGGSG-CTSQKWAQCGGIGFSGCTTCVSGTTC 288
           |  |:|  : |||:  | :     ||  |   |||| :||||
Db     330 ACACTTCGAGAAACGGGTTCCAAATGCACGTTGC--CTTCCAGTTC 373


RESULT 14
US-10-176-757-525
; Sequence 525, Application US/10176757
; Publication No. US20030022297A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C86
;  CURRENT APPLICATION NUMBER: US/10/176,757
;  CURRENT FILING DATE:  2002-06-20
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
;   LENGTH: 4440
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-176-757-525

  Query Match             9.3%;  Score 154;  DB 9;  Length 4440;
  Best Local Similarity   24.8%;  Pred. No. 0.001;
  Matches   71;  Conservative   19;  Mismatches  114;  Indels   82;  Gaps   14;

Qy       9 TTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSDFNVQS 68
           |    |       |:|:|  |     |   ||           ||       :    :
Db     164 TGTGTATGCATGCATGTGTGTG----CGTGCA----------CACGTG----TGTGTTT 204

Qy      69 GCNGGSAYSCADQTPWAVNDNLAYGF--AATSIAGGSESSWCCACYALTFTSGPVAGKTM 126
           ||  |:|  ||  |  |       |     |    |: :   |||   | |:|  ||
Db     205 GCGTGTA--CATGTGCATGTGTGTGTTGTGTGCCTGTGTGTGCAC--ATGTAGAAAGAAA 260

Qy     127 VVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQ---FGGLPGAQYGGISSRDQCDSFPA 183
             :| ||   |:         || |    | :     ||  ||  |  :    |
Db     261 GTGATGTGTGGGA-------ATGGAGAATTCCAACCCAGGAGGA--GACTGTGCCTG--- 308

Qy     184 PLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNGGTG 243
           ||                                :||                : | ||||
Db     309 --GGGC----------------------------TGC--------CACGAGAAGTGGTG 329

Qy     244 TPTSTAPGSGQTSPGGGSG-CTSQKWAQCGGIGFSGCTTCVSGTTC 288
           |  |:|  : |||:  | :     ||  |   |||| :||||
Db     330 ACACTTCGAGAAACGGGTTCCAAATGCACGTTGC--CTTCCAGTTC 373


RESULT 15
US-10-176-913-525
; Sequence 525, Application US/10176913
; Publication No. US20030022298A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C66
;  CURRENT APPLICATION NUMBER: US/10/176,913
;  CURRENT FILING DATE:  2002-06-20
;  Prior Application removed - See file Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
;   LENGTH: 4440
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-176-913-525

  Query Match             9.3%;  Score 154;  DB 9;  Length 4440;
  Best Local Similarity   24.8%;  Pred. No. 0.001;
  Matches   71;  Conservative   19;  Mismatches  114;  Indels   82;  Gaps   14;

Qy       9 TTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSDFNVQS 68
           |    |       |:|:|  |     |   ||           ||       :    :
Db     164 TGTGTATGCATGCATGTGTGTG----CGTGCA----------CACGTG----TGTGTTT 204

Qy      69 GCNGGSAYSCADQTPWAVNDNLAYGF--AATSIAGGSESSWCCACYALTFTSGPVAGKTM 126
           ||  |:|  ||  |  |       |     |    |: :   |||   | |:|  ||
Db     205 GCGTGTA--CATGTGCATGTGTGTGTTGTGTGCCTGTGTGTGCAC--ATGTAGAAAGAAA 260

Qy     127 VVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQ---FGGLPGAQYGGISSRDQCDSFPA 183
             :| ||   |:         || |    | :     ||  ||  |  :    |
Db     261 GTGATGTGTGGGA-------ATGGAGAATTCCAACCCAGGAGGA--GACTGTGCCTG--- 308

Qy     184 PLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNGGTG 243
           ||                                :||                : | ||||
Db     309 --GGGC----------------------------TGC--------CACGAGAAGTGGTG 329

Qy     244 TPTSTAPGSGQTSPGGGSG-CTSQKWAQCGGIGFSGCTTCVSGTTC 288
           |  |:|  : |||:  | :     ||  |   |||| :||||
Db     330 ACACTTCGAGAAACGGGTTCCAAATGCACGTTGC--CTTCCAGTTC 373


Search completed: February 24, 2003, 11:50:02
Job time : 17 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 24, 2003, 11:49:49 ; Search time 18 Seconds
(without alignments)
1596.900 Million cell updates/sec

Title: US-10-007-521-12
Perfect score: 1659
Sequence: 1 MRSTPVLRTTLAAALPLVAS.....TTCVSGTTCQKLNDYYSQCL 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	740.5	44.6	229	2	JC7308	cellulase (EC 3.2.
2	507	30.6	511	2	S10527	endoglucanase B pr
3	462	27.8	393	2	S59499	cellulase egl1 - s
4	221	13.3	242	2	S60143	cellulase (EC 3.2.
5	219.5	13.2	513	1	EUTQI	cellulose 1,4-beta
6	219.5	13.2	513	2	S45380	cellulose 1,4-beta
7	215.5	13.0	513	2	S11439	cellulose 1,4-beta
8	214	12.9	459	2	A25928	cellulase (EC 3.2.
9	203	12.2	463	2	A48375	cellulase (EC 3.2.
10	190	11.5	516	2	S33164	cellulose 1,4-beta
11	189	11.4	464	2	JC7143	endoglucanase I -
12	188.5	11.4	302	2	S71334	acetyl xylan ester
13	185	11.2	525	1	S38794	cellulose 1,4-beta
14	184	11.1	516	2	JS0083	cellulose 1,4-beta
15	183	11.0	537	1	JU0150	cellulose 1,4-beta
16	181	10.9	516	2	S42093	cellulose 1,4-beta
17	172	10.4	320	2	JC1311	cell protein precu
18	170.5	10.3	510	2	S41943	cellulose 1,4-beta
19	170.5	10.3	511	2	S44716	cellulose 1,4-beta
20	161	9.7	504	1	B48939	cellulose 1,4-beta
21	156	9.4	856	2	T00349	Avicelase III - As
22	150.5	9.1	860	2	S43846	xylanase B - rumen
23	143.5	8.6	388	1	JC5461	cellulase (EC 3.2.
24	143.5	8.6	388	1	S43920	cellulase (EC 3.2.
25	139.5	8.4	540	2	S41942	cellulose 1,4-beta
26	137	8.3	438	1	S70602	cellulose 1,4-beta
27	137	8.3	839	2	F75518	hypothetical prote
28	129	7.8	418	1	S28372	cellulase (EC 3.2.
29	127	7.7	429	1	JC5861	endo-1,4-beta-xyla
30	125.5	7.6	316	1	A38743	loricrin - human
31	125.5	7.6	410	1	S68153	cellulase (EC 3.2.
32	124	7.5	471	1	A26160	cellulose 1,4-beta
33	124	7.5	471	1	A38979	cellulose 1,4-beta
34	122	7.4	182	2	A36686	ultra-high-sulfur
35	120.5	7.3	303	2	H70716	hypothetical prote
36	120	7.2	5376	2	T42215	zonadhesin - mouse
37	118.5	7.1	453	2	A41640	vestigial protein
38	118.5	7.1	749	2	I38488	trophinin - human
39	116.5	7.0	584	2	G70804	hypothetical glyci
40	115.5	7.0	191	2	I46412	keratin KAP5.4 - s
41	114	6.9	461	2	T51044	related to spore c
42	114	6.9	1804	2	H96597	hypothetical prote
43	113.5	6.8	1381	2	E70806	hypothetical glyci
44	113.5	6.8	2090	2	S26058	probable transform
45	111.5	6.7	1032	2	T34433	hypothetical prote

ALIGNMENTS

RESULT 1
JC7308
cellulase (EC 3.2.1.4) - Scopulariopsis brevicaulis
N;Alternate names: endoglucanase I
C;Species: Scopulariopsis brevicaulis
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C;Accession: JC7308; PC7087
R;Nakatani, F.; Kawaguchi, T.; Takada, G.; Sumitani, J.; Moriyama, Y.; Arai, M.
Biosci. Biotechnol. Biochem. 64, 1238-1246, 2000
A;Title: Cloning and sequencing of an endoglucanase gene from Scopulariopsis brevicau
A;Reference number: JC7308
A;Accession: JC7308
A;Molecule type: DNA
A;Residues: 1-229 <NAK>
A;Experimental source: strain TOF-1212
A;Accession: PC7087
A;Molecule type: protein
A;Residues: 21-37;149-164 <NA2>
C;Genetics:
A;Gene: egI
A;Introns: 147/3
C;Keywords: glycosidase; hydrolase

Query Match 44.6%; Score 740.5; DB 2; Length 229;
Best Local Similarity 59.8%; Pred. No. 2.6e-45;
Matches 134; Conservative 32; Mismatches 49; Indels 9; Gaps 5;

```
Qy      10 TLAAALPLV---ASAASGSGQSTRYWDCCKPSCAWPGKAAVSQ-PVYACDANFQRLSDFN 65
           || |  |||   || |||:| :|||||||||||:|| || :|| |   || |   | |
Db       6 TLLALTPLVLPAASQASGTGTTTRYWDCCKPSCSWPDKAPLSQGPPMTCDINDNPLDDGG 65

Qy      66 V-QSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAG 123
           : :|||  || || |:  :||||:| ||||:|| :| | :|| |||||| | ||:| |:|
Db      66 LTESGCEPGGGAYMCSSHSPWAVDDELAYGWAAVNIGGQTESDWCCACYELEFTTGAVSG 125

Qy     124 KTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLP---GAQYGGISSRDQCDS 180
           | |:||:|:||||||:| |||||||||||||||: |:|  |   | :|||: :|  |||
Db     126 KKMIVQATNTGGDLGNNHFDIAMPGGGVGIFNGCTDQWGSPPNGWGDRYGGVHTRADCDS 185

Qy     181 FPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRN 224
           ||  || ||:|||||   |||  :|:::|:||||:| :| |:|:
Db     186 FPEALKAGCEWRFDWFGGTDNPDVSFREVECPAELVQKSQCQRS 229
```

RESULT 2
S10527
endoglucanase B precursor - Pseudomonas fluorescens
C;Species: Pseudomonas fluorescens
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
C;Accession: S10527
R;Gilbert, H.J.; Hall, J.; Hazlewood, G.P.; Ferreira, L.M.A.
Mol. Microbiol. 4, 759-767, 1990

```
A;Title: The N-terminal region of an endoglucanase from Pseudomonas fluorescens subspeci
A;Reference number: S10527; MUID:90355836; PMID:2117693
A;Accession: S10527
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-511 <GIL>
A;Cross-references: EMBL:X52615; NID:g45497; PIDN:CAA36844.1; PID:g45498
C;Superfamily: bacterial cellulose-binding domain homology; glycosidase GWGW domain homo
F;31-128/Domain: bacterial cellulose-binding domain homology <BCB>
F;180-217/Domain: glycosidase GWGW domain homology <GWG>
F;32-127/Disulfide bonds: #status predicted

  Query Match             30.6%;  Score 507;  DB 2;  Length 511;
  Best Local Similarity   39.6%;  Pred. No. 1.5e-28;
  Matches  107;  Conservative   38;  Mismatches   73;  Indels   52;  Gaps    7;

Qy       2 RSTPVLRTTLAAALPLVASAASG--------SGQSTRYWDCCKPSCAWPGKA-AVSQPVY 52
           ||:    ::|::|     ||: |         :| :|||||||||| | |     ::  |:
Db     237 RSSVASSSSLSSATSSSASSVSSVPPIDGGCNGYATRYWDCCKPHCGWSANVPSLVSPLQ 296

Qy      53 ACDANFQRLSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACY 112
           :| ||  |||| :| | |:||  | | |: |:||:  ||||:|||     |    |  ||
Db     297 SCSANNTRLSDVSVGSSCDGGGGYMCWDKIPFAVSPTLAYGYAAT-----SSGDVCGRCY 351

Qy     113 ALTFT-----------SGPVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQF 161
           | ||           |  :|||||:||:|: | |:   |||| :|||||| || ||:|:
Db     352 QLQFTGSSYNAPGDPGSAALAGKTMIVQATNIGYDVSGGQFDILVPGGGVGAFNACSAQW 411

Qy     162 G---GLPGAQYGGISSR------------------DQCDSF-----PAPLKPGCQWRFD 194
           |      ||||||  :                  ::|||        |: || |  :
Db     412 GVSNAELGAQYGGFLAACKQQLGYNASLSQYKSCVLNRCDSVFGSRGLTQLQQGCTWFAE 471

Qy     195 WFQNADNPTFTFQQVQCPAEIVARSGCKRN 224
           ||: ||||:  :::| ||||:  |||  |:
Db     472 WFEAADNPSLKYKEVPCPAELTTRSGMNRS 501


RESULT 3
S59499
cellulase egl1 - smut fungus (Ustilago maydis)
C;Species: Ustilago maydis (corn smut)
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 28-May-1999
C;Accession: S59499
R;Schauwecker, F.; Wanner, G.; Kahmann, R.
Biol. Chem. Hoppe-Seyler 376, 617-625, 1995
A;Title: Filament-specific expression of a cellulase gene in the dimorphic fungus Ustila
A;Reference number: S59499; MUID:96145728; PMID:8590631
A;Accession: S59499
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-393 <SCH>
A;Cross-references: GB:S81598; NID:g1478378; PIDN:AAB36147.1; PID:g1478379

  Query Match             27.8%;  Score 462;  DB 2;  Length 393;
  Best Local Similarity   37.7%;  Pred. No. 1.8e-25;
  Matches  117;  Conservative   41;  Mismatches  106;  Indels   46;  Gaps   16;

Qy      11 LAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRL----SDFNV 66
           |: :| ||      :| :|||||||  | :| ||| |  || || |:   |     | :
Db      12 LSLSLSLVHLDGVRAGMATRYWDCCLASASWEGKAPVYAPVDACKADGVTLIDSKKDPSG 71

Qy      67 QSGCNGGSAYSCADQTPW--AVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGK 124
           |||||||: : |:   |:    :  ||:|| |     | ||   |||:   |      ||
Db      72 QSGCNGGNKFMCSCMQPFDDETDPTLAFGFGA--FTTGQESDTDCACFYAEF-EHDAQGK 128

Qy     125 TM-----VVQSTSTGGDLGSNQFDIAMPGGGVGIF-NGCSSQFG---GLPGAQYGGISSR 175
           |     : | |: |||: |  ||  :||||:| |  || :|:|    | | ||||: |
Db     129 AMKRNKLIFQVTNVGGDVQSQNFDFQIPGGGLGAFPKGCPAQWGVEASLWGDQYGGVKSA 188

Qy     176 DQCDSFPAPLKPGCQWRF-DWFQNADNPTF--TFQQVQCPAEIVARSGCKRNDDSSFPVF 232
           :|   | ||: ||:||| :|    |||    : ::|:||  :: ||||:| ||::
Db     189 TECSKLPKPLQEGCKWRFSEW---GDNPVLKGSPKRVKCPKSLIDRSGCQRKDDNT---I 242
```

```
Qy     233 TPPSG--GNGGTGTPTS--------TAPG-SGQTSPGG--GSGCTSQKWAQCGGIGFSGC 279
           :| ||   :  | |           | |  |::: ||  |||  |      ||   ||
Db     243 SPYSGKVDSANTAAPAQYKRDRSVCLAGGKKGKSAAGGVDGSGDAS------GGADASGA 296

Qy     280 TTCVSGTTCQ 289
             |:  |
Db     297 GGAAEGSQGQ 306


RESULT 4
S60143
cellulase (EC 3.2.1.4) V precursor - fungus (Trichoderma reesei) (strain QM9414)
N;Alternate names: endo-1,4-beta-glucanase V
C;Species: Trichoderma reesei
A;Variety: strain QM9414
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 17-Mar-1999
C;Accession: S60143; S49043
R;Saloheimo, A.; Henrissat, B.; Hoffren, A.; Teleman, O.; Penttila, M.
submitted to the EMBL Data Library, May 1994
A;Description: A novel small endoglucanase gene egl5 from Trichoderma reesei isolated
A;Reference number: S60143
A;Accession: S60143
A;Molecule type: DNA
A;Residues: 1-242 <SAL>
A;Cross-references: EMBL:Z33381; NID:g485863; PID:g485864
R;Saloheimo, A.; Henrissat, B.; Hoffren, A.M.; Teleman, O.; Penttilae, M.
Mol. Microbiol. 13, 219-228, 1994
A;Title: A novel, small endoglucanase gene, egl5, from Trichoderma reesei isolated by
A;Reference number: S49043; MUID:95075308; PMID:7984103
A;Accession: S49043
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 206-241 <SAW>
A;Cross-references: EMBL:Z33381
C;Genetics:
A;Gene: egl5
A;Introns: 23/3; 45/3
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A;Pathway: cellulose degradation
C;Superfamily: fungal cellulose-binding domain homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-242/Product: cellulase #status predicted <MAT>
F;210-241/Domain: fungal cellulose-binding domain homology <FCB>

  Query Match             13.3%;  Score 221;  DB 2;  Length 242;
  Best Local Similarity   24.7%;  Pred. No. 1.1e-08;
  Matches   78;  Conservative   39;  Mismatches  101;  Indels   98;  Gaps   13;

Qy       7 LRTTLAAALPLVASAASGSGQSTRYWD-------CCKPSCAWPGKAAVSQPVYACDANFQ 59
           :: ||      :| : ::    :|||:|       |   | |:| :  :   ||   |  |
Db       1 MKATLVLGSLIVGAVSAYKATTTRYYDGQEGACGCGSSSGAFPWQLGIGNGVYTA-AGSQ 59

Qy      60 RLSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFT-- 117
           | |    | |  |                                    |  || || |
Db      60 ALFDTAGASWCGAG---------------------------------CGKCYQLTSTGQ 85

Qy     118 -------SGPVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYG 170
                 :|  ||:::::|  |:   : |: |:   :  ||                  ||
Db      86 APCSSCGTGGAAGQSIIVMVTNLCPNNGNAQWCPVV--GGTN--------------QY- 127

Qy     171 GISSRDQCDSFPAPLKPGCQWRFD-WFQN---ADNPTFTFQQVQCPAEIVARSG---CKR 223
                         |  : ||    ||    |: : || :  :  |   |
Db     128 ----------------GYSYHFDIMAQNEIFGDNVVVDFEPIACPGQAASDWGTCLCVG 170

Qy     224 NDDSSFPVFTPPSGGNGGTGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCV 283
             ::  |  ||  | : |:  | |: | :  : | ||   |   : |||| |::| |||
Db     171 QQETD-P--TPVLGNDTGSTPPGSSPPATSSSPPSGGGQQT--LYGQCGGAGWTGPTTCQ 225

Qy     284 SGTTCQKLNDYYSQCL 299
           :  ||:  | :|||||
Db     226 APGTCKVQNQWYSQCL 241
```

```
RESULT 5
EUTQI
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) I precursor - fungus (Trichoderma reesei
N;Alternate names: exo-cellobiohydrolase I
C;Species: Trichoderma reesei
C;Date: 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change 16-Jul-1999
C;Accession: A00902
R;Shoemaker, S.; Schweickart, V.; Ladner, M.; Gelfand, D.; Kwok, S.; Myambo, K.; Innis,
Bio/Technology 1, 691-696, 1983
A;Title: Molecular cloning of exo-cellobiohydrolase I derived from Trichoderma reesei st
A;Reference number: A00902
A;Accession: A00902
A;Molecule type: DNA
A;Residues: 1-513 <SHO>
A;Experimental source: strain L27
C;Comment: This is the most abundantly produced cellulase in this filamentous fungus; it
C;Genetics:
A;Gene: CBH1
A;Introns: 154/2; 386/3
C;Function:
A;Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in cellulose to r
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homo
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-513/Product: cellulose 1,4-beta-cellobiosidase I #status predicted <MAT>
F;482-513/Domain: fungal cellulose-binding domain homology <FCB>

  Query Match             13.2%;  Score 219.5;  DB 1;  Length 513;
  Best Local Similarity   26.9%;  Pred. No. 2.7e-08;
  Matches   91;  Conservative   36;  Mismatches  122;  Indels   89;  Gaps   22;

Qy      16 PLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPV--YACDANFQRLSDFNVQSGCNGG 73
           |   :| :| |    :  ||     |    ::|: :  : |    | : :     || |
Db     211 PSSNNANTGIG---GHGSCCSEMDIWEAN-SISEALTPHPCTTVGQEICE---GDGC--G 261

Qy      74 SAYS-------C----ADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVA 122
             ||       |     | |: : :   ||       | |         : |  |
Db     262 GTYSDNRYGGTCDPDGCDWNPYRLGNTSFYG------PGSS--------FTLDTTK---- 303

Qy     123 GKTMVVQSTSTGGDLG----SNQFDIAMPGGGVGIFNG-------CS---SQFGGLPGAQ 168
           |  ||    | | :      |      |   :| ::|       |:   ::|||   :
Db     304 -KLTVVTQFETSGAINRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSSFSD 362

Qy     169 YGGISSRDQCDSFPAPLKPGCQWRFDWFQN---ADNPTFTFQQVQCPAEIVARSGCKRND 225
           ||::   :  |    |     |  |::|    |:   | :    | : |  |  :
Db     363 KGGLTQFKKATSGGMVLVMSL-WD-DYYANMLWLDSTYPTNETSSTPGAV--RGSCSTSS 418

Qy     226 ------DSSFP---------VFTP------PSGGN---GGTGTPTSTAPGSGQTSPGGGS 261
               :|  |          | |      |||||   |  || |:  |    |: |
Db     419 GVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSGGNPPGGNRGTTTTRRPA---TTTGSSP 475

Qy     262 GCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           | |   : ||||||:|| | | ||||| || ||||||
Db     476 GPTQSHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL 513


RESULT 6
S45380
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - fungus (Trichoderma koningii)
C;Species: Trichoderma koningii
C;Date: 20-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C;Accession: S45380
R;Wey, T.T.; Hseu, T.H.; Huang, L.
Curr. Microbiol. 28, 31-39, 1994
A;Title: Molecular cloning and sequence analysis of the cellobiohydrolase I gene from Tr
A;Reference number: S45380; MUID:94100788; PMID:7764306
A;Accession: S45380
A;Molecule type: DNA
A;Residues: 1-513 <WEY>
A;Cross-references: EMBL:X69976; NID:g457422; PIDN:CAA49596.1; PID:g457423
C;Genetics:
A;Introns: 154/2; 386/3
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain h
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;482-513/Domain: fungal cellulose-binding domain homology <FCB>

  Query Match             13.2%;  Score 219.5;  DB 2;  Length 513;
  Best Local Similarity   26.9%;  Pred. No. 2.7e-08;
  Matches   91;  Conservative   36;  Mismatches  122;  Indels   89;  Gaps   22;

Qy      16 PLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPV--YACDANFQRLSDFNVQSGCNGG 73
           |   :| :| |    :  ||     |    ::|: :  : |    | : :     || |
Db     211 PSSNNANTGIG---GHGSCCSEMDIWEAN-SISEALTPHPCTTVGQEICE---GDGC--G 261

Qy      74 SAYS-------C----ADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVA 122
             ||       |     | |: : :   ||       | |         : |  |
Db     262 GTYSDNRYGGTCDPDGCDWNPYRLGNTSFYG------PGSS--------FTLDTTK---- 303

Qy     123 GKTMVVQSTSTGGDLG----SNQFDIAMPGGGVGIFNG-------CS---SQFGGLPGAQ 168
           |  ||    | | :      |      |   :| ::|       |:   ::|||   :
Db     304 -KLTVVTQFETSGAINRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSSFSD 362

Qy     169 YGGISSRDQCDSFPAPLKPGCQWRFDWFQN---ADNPTFTFQQVQCPAEIVARSGCKRND 225
           ||::   :  |    |     |  |::|    |:   | :    | : |  |  :
Db     363 KGGLTQFKKATSGGMVLVMSL-WD-DYYANMLWLDSTYPTNETSSTPGAV--RGSCSTSS 418

Qy     226 ------DSSFP---------VFTP------PSGGN---GGTGTPTSTAPGSGQTSPGGGS 261
               :|  |          | |      |||||   |  || |:  |    |: |
Db     419 GVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSGGNPPGGNRGTTTTRRPA---TTTGSSP 475

Qy     262 GCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           | |   : ||||||:|| | | ||||| || ||||||
Db     476 GPTQSHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL 513


RESULT 7
S11439
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - fungus (Trichoderma viride)
C;Species: Trichoderma viride
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C;Accession: S11439
R;Cheng, C.; Tsukagoshi, N.; Udaka, S.
Nucleic Acids Res. 18, 5559, 1990
A;Title: Nucleotide sequence of the cellobiohydrolase gene from Trichoderma viride.
A;Reference number: S11439; MUID:91016856; PMID:2216737
A;Accession: S11439
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-513 <CHE>
A;Cross-references: EMBL:X53931; NID:g5196; PIDN:CAA37878.1; PID:g295937
C;Genetics:
A;Introns: 154/2; 386/3
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain h
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;482-513/Domain: fungal cellulose-binding domain homology <FCB>

  Query Match             13.0%;  Score 215.5;  DB 2;  Length 513;
  Best Local Similarity   26.2%;  Pred. No. 5.2e-08;
  Matches   86;  Conservative   36;  Mismatches  137;  Indels   69;  Gaps   16;

Qy      16 PLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPV--YACDANFQRLSDFNVQSGCNGG 73
           |   :| :| |    :  ||     |    ::|: :  : |    | : : :   |  |
Db     211 PSSNNANTGIG---GHGSCCSEMDIWEAN-SISEALTPHPCTTVGQEICEGDSCGGTYSG 266

Qy      74 SAY--SC----ADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMV 127
             | :|     |  |: : :   ||       | |         : | |      |  |
Db     267 DRYGGTCDPDGCDWNPYRLGNTSFYG------PGSS--------FTLDTTK-----KLTV 307

Qy     128 VQSTSTGGDLG----SNQFDIAMPGGGVGIFNGCS----------SQFGGLPGAQYGGIS 173
           |    | | :      |      |   :| ::| |           ::|||   :  ||::
Db     308 VTQFETSGAINRYYVQNGVTFQQPNAELGDYSGNSLDDDYCAAEEAEFGGSSFSDKGGLT 367

Qy     174 SRDQCDSFPAPLKPGCQWRFDWFQN---ADNPTFTFQQVQCPAEI----VARSGCKRNDD 226
             : |    |     | |::  |    |:   | :    |  :       ||     :
```

```
Db    368 QFKKATSGGMVLVMSL-WD-DYYANMLWLDSTYPTDETSSTPGAVRGSSSTSSGVPAQLE 425

Qy    227 SSFP---------VFTP------PSGGNGGTGTPTSTAPGSGQTSPGGGSGCTSQKWAQC 271
          |: |         | |      |||||   | |  |      || |   | |   : ||
Db    426 SNSPNAKVVYSNIKFGPIGSTGNPSGGNPPGGNPPGTTTPRPATSTGSSPGPTQTHYGQC 485

Qy    272 GGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
          ||||: | | | ||:||| || ||||||
Db    486 GGIGYIGPTVCASGSTCQVLNPYYSQCL 513


RESULT 8
A25928
cellulase (EC 3.2.1.4) A precursor - fungus (Trichoderma reesei)
N;Alternate names: endo-1,4-beta-glucanase; endoglucanase I
C;Species: Trichoderma reesei
C;Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 22-Jun-1999
C;Accession: A25928; A25565
R;Van Arsdell, J.N.; Kwok, S.; Schweickart, V.L.; Ladner, M.B.; Gelfand, D.H.; Innis, M.
Bio/Technology 5, 60-64, 1987
A;Title: Cloning, characterization, and expression in Saccharomyces cerevisiae of endogl
A;Reference number: A25928
A;Accession: A25928
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-459 <VAN>
A;Experimental source: strain L27
R;Penttilae, M.; Lehtovaara, P.; Nevalainen, H.; Bhikhabhai, R.; Knowles, J.
Gene 45, 253-263, 1986
A;Title: Homology between cellulase genes of Trichoderma reesei: complete nucleotide seq
A;Reference number: A25565; MUID:87106822; PMID:2948877
A;Accession: A25565
A;Molecule type: mRNA
A;Residues: 1-459 <PEN>
A;Cross-references: GB:M15665; NID:g170546; PIDN:AAA34212.1; PID:g170547
C;Genetics:
A;Gene: egl1
A;Introns: 257/2; 454/1
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A;Pathway: cellulose degradation
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homo
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-459/Product: cellulase A #status predicted <MAT>
F;428-459/Domain: fungal cellulose-binding domain homology <FCB>

  Query Match             12.9%;  Score 214;  DB 2;  Length 459;
  Best Local Similarity   27.0%;  Pred. No. 6e-08;
  Matches  108;  Conservative   31;  Mismatches  101;  Indels  160;  Gaps   27;

Qy     19 ASAASGSGQS---TRYWDCCKPSCAWPGKAAVSQPVYACDA---------NFQRLSDFNV 66
          ||  : || |    :|     || :  | ::||  :|  |:         | | || |:|
Db    101 ASGVTTSGSSLTMNQY----MPSSS-GGYSSVSPRLYLLDSDGEYVMLKLNGQELS-FDV 154

Qy     67 Q---------------------------SGCNGGSAYSCADQTP---WAVNDNLAYGFA 95
                                       :| | || | |  | |   |  | |
Db    155 DLSALPCGENGSLYLSQMDENGGANQYNTAGANYGSGY-CDAQCPVQTWR-NGTL----- 207

Qy     96 ATSIAGGSESSWCC----------------------AC---------YALTFTS--GPV 121
              |  :||                          ||         |   : |  ||
Db    208 -----NTSHQGFCCNEMDILEGNSRANALTPHSCTATACDSAGCGFNPYGSGYKSYYGP- 261

Qy    122 AGKTMVVQST------------STGGDLGS-------NQFDI--AMPGGGVGIFNGCSSQ 160
           | |:    |            |  |:| |       |  ||  | |||     : | |
Db    262 -GDTVDTSKTFTIITQFNTDNGSPSGNLVSITRKYQQNGVDIPSAQPGGDT--ISSCPS- 317

Qy    161 FGGLPGAQYGGISSRDQCDSFPAPLKPGCQWRFD-WFQNAD--NPTFTFQQVQC------ 211
                : |||:::  :       |  |    |  |  |:   |   :     |
Db    318 -----ASAYGGLATMGKA------LSSGMVLVFSIWNDNSQYMNWLDSGNAGPCSSTEGN 366

Qy    212 PAEIVARSGCKRNDDSSFPVFTPPSGGNGGTGTPTSTA----PGSGQT--------SPGG 259
          |: |:|      |: ::  ||:    |: |: |  |||    | |  |        :

Db    367 PSNILA------NNPNTHVVFSNIRWGDIGS-TTNSTAPPPPPASSTTFSTTRRSSTTSS 419

Qy    260 GSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
             ||    | |||||||:||| || ||||||  ||||||||
Db    420 SPSCTQTHWGQCGGIGYSGCKTCTSGTTCQYSNDYYSQCL 459


RESULT 9
A48375
cellulase (EC 3.2.1.4) - fungus (Trichoderma longibrachiatum)
N;Alternate names: endo-1,4-beta-glucanase; endoglucanase I homolog EGL1
C;Species: Trichoderma longibrachiatum
C;Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 22-Jun-1999
C;Accession: A48375; S28521
R;Gonzalez, R.; Ramon, D.; Perez-Gonzalez, J.A.
Appl. Microbiol. Biotechnol. 38, 370-375, 1992
A;Title: Cloning, sequence analysis and yeast expression of the egl1 gene from Tricho
A;Reference number: A48375; MUID:93159747; PMID:1369161
A;Contents: CECT 2606
A;Accession: A48375
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-463 <GON>
A;Cross-references: EMBL:X60652; NID:g5181; PIDN:CAA43059.1; PID:g5182
A;Note: sequence extracted from NCBI backbone (NCBIN:125157, NCBIP:125158)
C;Genetics:
A;Introns: 257/2; 458/1
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A;Pathway: cellulose degradation
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain h
C;Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F;432-463/Domain: fungal cellulose-binding domain homology <FCB>

  Query Match             12.2%;  Score 203;  DB 2;  Length 463;
  Best Local Similarity   28.4%;  Pred. No. 3.6e-07;
  Matches   86;  Conservative   22;  Mismatches  101;  Indels   94;  Gaps   18;

Qy     68 SGCNGGSAYSCADQTP---WAVNDNL---AYGFAAT--SIAGGSESSWCCACYALTFTSG 119
          :| | || | |  | |   |  | |  |      ||       |  |:  :     :: | |:
Db    184 AGANYGSGY-CDAQCPVQTWR-NGTLNTSGQGFCCNEMDILEGNSRANALTPHSCTATAC 241

Qy    120 PVAG----------------------KTMVV------QSTSTGGDLGS-------NQFD 143
            ||                      ||  :      : |  |:| |       | |
Db    242 DSAGCGFNPYGSGYPNYFGPGDTVDTSKTFTIITQFNTDNGSPSGNLVSITRKYRQNGVD 301

Qy    144 I--AMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCDSFPAPLKPGCQWRFD-WFQNAD 200
          |  | |||     : | |       : |||:::  :       |  |    |  |  |:
Db    302 IPSAKPGGDT--ISSCPS------ASAYGGLATMGKA------LSSGMVLVFSIWNDNSQ 347

Qy    201 -----------------NPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGN--- 239
                           ||:        |  |  |  :  |  |    |  :|||
Db    348 YMNWLDSGRAGPCSSTEGNPSNIL--ANNPGTHVVYSNIRWGDIGS---TTNSTGGNPPP 402

Qy    240 ---GGTGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYS 296
             : |  ||   |  ||      ||   | ||||||::|| || ||||||  |||||
Db    403 PPPPASSTTFSTTRRSSTTS--SSPSCTQTHWGQCGGIGYTGCKTCTSGTTCQYGNDYYS 460

Qy    297 QCL 299
          |||
Db    461 QCL 463


RESULT 10
S33164
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - basidiomycete (Phanerochaete chryso
C;Species: Phanerochaete chrysosporium
C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C;Accession: S33164
R;Sims, P.F.G.; Soares-Felipe, S.M.; Gent, M.E.; Tempelaars, C.; Wang, Q.; Broda, P.
submitted to the EMBL Data Library, April 1993
A;Description: Differential expression of multiple exo-cellobiohydrolase I-like genes
A;Reference number: S33164
```

```
A;Accession: S33164
A;Molecule type: DNA
A;Residues: 1-516 <SIM>
A;Cross-references: EMBL:Z22528; NID:g296026; PIDN:CAA80253.1; PID:g296027
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homo
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;485-516/Domain: fungal cellulose-binding domain homology <FCB>

  Query Match             11.5%;  Score 190;  DB 2;  Length 516;
  Best Local Similarity   24.4%;  Pred. No. 3.3e-06;
  Matches   88;  Conservative   34;  Mismatches  101;  Indels  138;  Gaps   18;

Qy      20 SAASGSGQSTRYWDCCKPSCAWPGK--AAVSQPVYACDANFQRLSDFNVQSGCNGG---- 73
           || :|:|    |  ||     |     ||   | : |          | |: |:|
Db     211 SANAGTG---NYGTCCTEMDIWEANNDAAAYTP-HPCTT--------NAQTRCSGSDCTR 258

Qy      74 -----------SAYSCADQT-----------PWAV------NDNLAYGFAATSIAGGS 103
                      :::   |||            |: |      ||  : |
Db     259 DTGLCDADGCDFNSFRMGDQTFLGKGLTVDTSKPFTVVTQFITNDGTSAG---------- 308

Qy     104 ESSWCCACYALTFTSGPVAGKTMVVQSTST---GGDLGSNQFDIAMPGGGVGIFNGCSSQ 160
                  ||          |:|::|    | || ::  |             | || |
Db     309 ---------TLTEIRRLYVQNGKVIQNSSVKIPGIDLVNSITD----------NFCSQQ 348

Qy     161 ---FGGLP-GAQYGGISSRDQC-----------DSFPAPLKPGCQWRFDWFQ-----NA 199
               ||     ||:||:    :           | : | :          |       |
Db     349 KTAFGDTNYFAQHGGLKQVGEALRTGMVLALSIWDDYAANML--------WLDSNYPTNK 400

Qy     200 DNPTFTFQQVQC------PAEIVARSGCKRNDDSSFPVFTPPSGGNGGT---------- 242
           |  |    :  |      ||:| |:|       ::: ||:    |:  |
Db     401 DPSTPGVARGTCATTSGVPAQIEAQS------PNAYVVFSNIKFGDLNTTYTGTVSSSSV 454

Qy     243 -----GTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQ 297
                | ||::  |  | |   :| |  :| ||||||::| ||| |  ||  || ||||
Db     455 SSSHSSTSTSSSHSSSSTPPTQPTGVTVPQWGQCGGIGYTGSTTCASPYTCHVLNPYYSQ 514

Qy     298 C 298
           |
Db     515 C 515


RESULT 11
JC7143
endoglucanase I - fungus (Trichoderma viride)
C;Species: Trichoderma viride
C;Date: 04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 09-Jun-2000
C;Accession: JC7143
R;Kwon, I.; Ekino, K.; Goto, M.; Furukawa, K.
Biosci. Biotechnol. Biochem. 63, 1714-1720, 1999
A;Title: Heterologous expression and characterization of endoglucanase I (EGI) from Tric
A;Reference number: JC7143; MUID:20052947; PMID:10586500
A;Accession: JC7143
A;Molecule type: mRNA
A;Residues: 1-464 <KWO>
A;Experimental source: HK-75
C;Comment: This protein is a fusion glycoprotein with catalytic and cellulose binding do
C;Genetics:
A;Gene: egl1
A;Introns: 257/2; 459/1
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homo
C;Keywords: fusion protein; glycoprotein
F;433-464/Domain: fungal cellulose-binding domain homology <FCB>

  Query Match             11.4%;  Score 189;  DB 2;  Length 464;
  Best Local Similarity   25.4%;  Pred. No. 3.5e-06;
  Matches  102;  Conservative   32;  Mismatches  108;  Indels  160;  Gaps   25;

Qy      21 AASG---SGQS---TRYWDCCKPSCAWPGKAAVSQPVYACDA---------NFQRLSDFN 65
           ||||   || |    :|     || :  | ::||  :|  |:         | | || |:
Db     100 AASGVQTSGASLTMNQY----MPSSS-GGYSSVSPRLYLLDSDGEYVLLKLNGQELS-FD 153

Qy      66 VQ--------------------------SGCNGGSAYSCADQTP---WAVNDNL---A 91
           |                           :| | || | |  | |   |  |  |
Db     154 VDLSALPCGENGSLYLSEMDETGGANQYNTAGANYGSGY-CDAQCPVQTWR-NGTLNTSH 211

Qy      92 YGFAAT--SIAGGSESSWCCACYALTFTSGPVAG------------------------K 124
             ||      |  |:  :     :: | |:   ||
Db     212 QGFCCNEMDILEGNSRANALTPHSCTATACDSAGCGFNPYGSGYPNYYGPGDTVDTSKVF 271

Qy     125 TMVVQ-STSTGGDLGS----------NQFDI--AMPGGGVGIFNGCSSQFGGLPGAQYGG 171
           |:: | :|  |   |:          |  ||  | |||     : | |       : |||
Db     272 TIITQFNTDDGSATGNLVSITRKYRQNGVDIPSAKPGGDT--ISSCPS------ASAYGG 323

Qy     172 ISSRDQCDSFPAPLKPGCQWRFD-WFQNAD-----------------NPTFTFQQVQCP 212
           :::  :        |  |    |  |  |:                    |||
Db     324 LTTMGKA------LSSGMVLVFSIWNDNSQYMNWLDSGSAGPCSSTEGNPT--------- 368

Qy     213 AEIVARS---------------GCKRNDDSSFPVFTPPSGGNGGTGTPTSTAPGSGQTSP 257
             |:| :               |   |   : |   ||      : |  ||   |  ||
Db     369 -NILANNPNTHVVFSNIRWGDIGSTTNSTGTTP---PPPPPPASSTTLSTTRRSSTTS- 423

Qy     258 GGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
                ||   : || |||::|| || :|||||  ||||||||
Db     424 -SSPSCTQTHYGQCNGIGYTGCKTCAAGTTCQYSNDYYSQCL 464


RESULT 12
S71334
acetyl xylan esterase precursor - fungus (Trichoderma reesei)
C;Species: Trichoderma reesei
C;Date: 23-Jul-1997 #sequence_revision 01-Aug-1997 #text_change 17-Mar-1999
C;Accession: S71334
R;Margolles-Clark, E.; Tenkanen, M.; Soederlund, H.; Penttilae, M.
Eur. J. Biochem. 237, 553-560, 1996
A;Title: Acetyl xylan esterase from Trichoderma reesei contains an active-site serine
A;Reference number: S71334; MUID:96235218; PMID:8647098
A;Accession: S71334
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-302 <MAR>
A;Cross-references: EMBL:Z69256; NID:g1431619; PID:e220701; PID:g1431620
C;Genetics:
A;Gene: axe1
C;Superfamily: fungal cellulose-binding domain homology
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-302/Product: acetyl xylan esterase #status predicted <MAT>
F;271-302/Domain: fungal cellulose-binding domain homology <FCB>

  Query Match             11.4%;  Score 188.5;  DB 2;  Length 302;
  Best Local Similarity   26.6%;  Pred. No. 2.5e-06;
  Matches   73;  Conservative   17;  Mismatches   97;  Indels   87;  Gaps    8;

Qy      92 YGFAATSI-----------------AGGSESSWCCACYALTFTSGPVAGKTMV------ 127
           || :|| :                 | | ::|     || :  :|  |    :
Db      50 YGSSATVVNLVIQAHPGTTSEAIVYPACGGQASCGGISYANSVVNGTNAAAAAINNFHNS 109

Qy     128 ---VQSTSTGGDLGSNQFDIAMPGG---GVGIFNGCSSQFGGLPGAQYGGISSRDQCDSF 181
              |    |   |:  || |: ||   | || |       |   |    |   |  :
Db     110 CPDTQLVLVGYSQGAQIFDNALCGGGDPGEGITNTAVPLTAGAVSAVKAAIFMGDPRNIH 169

Qy     182 PAPLKPGCQWRFDWFQNADNPTFTFQQVQ-------CPAEIVARSGCKR----------- 223
           |   |               | | |           ||:    :| |
Db     170 GLPYNVG-------------TCTTQGFDARPAGFVCPSASKIKSYCDAADPYCCTGNDP 215

Qy     224 -----------------NDDSSFPVFTPPSGGNGGTGTPTSTAPGSGQTSPGGGSGCTS 265
                           |   |      || ||   |  ||||  ||   |||     |
Db     216 NVHQGYGQEYGQQALAFINSQLSSGGSQPPGGGPTSTSRPTSTRTGS---SPGP----TQ 268

Qy     266 QKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           | |||| |::| | | |||||| :: :|||||
Db     269 THWGQCGGQGWTGPTQCESGTTCQVISQWYSQCL 302


RESULT 13
S38794
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - imperfect fungus (Humicola grisea)
```

```
N;Alternate names: beta-glucancellobiohydrolase; exoglucanase
C;Species: Humicola grisea var. thermoidea
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S38794; S08240; A45869
R;Radford, A.
submitted to the EMBL Data Library, June 1991
A;Reference number: S38794
A;Accession: S38794
A;Molecule type: DNA
A;Residues: 1-525 <RAD>
A;Cross-references: EMBL:X17258; NID:g2760; PIDN:CAA35159.1; PID:g2761
A;Note: this is a revision to the sequence from reference S08240
R;de Oliviera Azevedo, M.; Radford, A.
Nucleic Acids Res. 18, 668, 1990
A;Title: Sequence of cbh-1 gene of Humicola grisea var. thermoidea.
A;Reference number: S08240; MUID:90175006; PMID:2308855
A;Accession: S08240
A;Molecule type: DNA
A;Residues: 1-299,'H',301-525 <DEO>
A;Cross-references: EMBL:X17258
A;Note: the authors translated the codon CAG for residue 87 as His
A;Note: this sequence has been revised in reference S38794
R;Azevedo, M.; de, O.; Felipe, M.S.S.; Astolfi-Filho, S.; Radford, A.
J. Gen. Microbiol. 136, 2569-2576, 1990
A;Title: Cloning, sequencing and homologies of the cbh-1 (exoglucanase) gene of Humicola
A;Reference number: A45869; MUID:91178527; PMID:2127803
A;Accession: A45869
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-20,'R',22-34,'K',36-86,'H',88-141,'V',143-157,'Y',159-237,'QQH',241-244,'I
A;Cross-references: GB:M64588; GB:X17258
A;Note: this sequence has been revised. See entry S08240
C;Genetics:
A;Gene: cbh-1
A;Introns: 138/1
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homo
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;494-525/Domain: fungal cellulose-binding domain homology <FCB>

  Query Match             11.2%;  Score 185;  DB 1;  Length 525;
  Best Local Similarity   24.9%;  Pred. No. 7.5e-06;
  Matches   97;  Conservative   33;  Mismatches  136;  Indels  124;  Gaps   23;

Qy      11 LAAALPLVASAASGSGQSTRYWDCCKPSCAWPG-KAAVSQPVYACDANFQRLSDF----- 64
           |  ||  |:  | |    :||          || ||        |||   |   |
Db     159 LNGALYFVSMDADGG--LSRY----------PGNKAGAKYGTGYCDAQCPRDIKFINGEA 206

Qy      65 NVQSGCN-------GGSAYS--CADQTPWAVNDNLAYGFA--ATSIAGGS--ESSWCCAC 111
           |::           |   |   |::   |  | |:|  |     :| | |  |   |
Db     207 NIEGWTGSTNDPNAGAGRYGTCCSEMDIWEAN-NMATAFTPHPCTIIGQSRCEGDSCGGT 265

Qy     112 YALTFTSG-----------------PVAGKTMVVQST-----------STGGDLGS-NQF 142
           |:    :|                  || | | :|           ||||   :|
Db     266 YSNERYAGVCDPDGCDFNSYRQGNKTFYGKGMTVDTTKKITVVTQFLKDANGDLGEIKRF 325

Qy     143 -----------DIAMPG--GGVGIFNGCSSQ---FGGLPG-AQYGGISSRDQCDSFPAPL 185
                      :  :||  |     : |  |   || :    : ||:    :  : |  |
Db     326 YVQDGKIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMGKALAGPMVL 385

Qy     186 KPGCQWRFDWFQNADNPTF---TFQQVQCPAEIVARSGCKRN---DDSSFPV-------- 231
                 |  :| |  :   ||     | :   : | :|      |  |
Db     386 VMSI-----WDDHASNMLWLDSTF-----PVDAAGKPGAERGACPTTSGVPAEVEAEAPN 435

Qy     232 ---------FTP--------PSGGNGGTG-----TPTSTAPGSGQTSPGGGSGCTSQKWA 269
                   | |         |  ||||        ||:|   :  |:    :|  : :|
Db     436 SNVVFSNIRFGPIGSTVAGLPGAGNGGNNGGNPPPPTTTTSSAPATTTTASAGPKAGRWQ 495

Qy     270 QCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           |||||||:| | |     | ||||:|||||
Db     496 QCGGIGFTGPTQCEEPYICTKLNDWYSQCL 525


RESULT 14
```

```
JS0083
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) I - basidiomycete (Phanerochaete chry
C;Species: Phanerochaete chrysosporium
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 04-Sep-1998
C;Accession: JS0083
R;Sims, P.; James, C.; Broda, P.
Gene 74, 411-422, 1988
A;Title: The identification, molecular cloning and characterisation of a gene from Ph
A;Reference number: JS0083; MUID:89232732; PMID:3246351
A;Accession: JS0083
A;Molecule type: DNA
A;Residues: 1-516 <SIM>
A;Experimental source: strain ME446
C;Comment: This protein is a component of the inducible cellulase complex.
C;Genetics:
A;Introns: 201/3; 511/1
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain h
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;485-516/Domain: fungal cellulose-binding domain homology <FCB>
F;81,92/Active site: Asp #status predicted

  Query Match             11.1%;  Score 184;  DB 2;  Length 516;
  Best Local Similarity   24.7%;  Pred. No. 8.7e-06;
  Matches   86;  Conservative   37;  Mismatches  113;  Indels  112;  Gaps   20;

Qy      20 SAASGSGQSTRYWDCCKPSCAWPGK--AAVSQPVYACDANFQRLSDFNVQSGCNGG---- 73
           || :|:|    |  ||     |     ||   | : |          | |: |:|
Db     211 SANAGTG---NYGTCCTEMDIWEANNDAAAYTP-HPCTT--------NAQTRCSGSDCTR 258

Qy      74 ------------SAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPV 121
                        :::   |||           |    :   :   :      :|   |
Db     259 DTGLCDADGCDFNSFRMGDQT----------FLGKGLTVDTSKPFTVVTQFIT-NDGTS 306

Qy     122 AG-----KTMVVQSTSTGGDLGSNQFDIAMPG-GGVGIF--NGCSSQ---FGGLP-GAQY 169
           ||     : : ||:    | :  |   : :||   |     | || |   ||     ||:
Db     307 AGTLTEIRRLYVQN----GKVIQNS-SVKIPGIDPVNSITDNFCSQQKTAFGDTNYFAQH 361

Qy     170 GGISSRDQC------------DSFPAPLKPGCQWRFDWFQ-----NADNPTFTFQQVQC- 211
           ||:    :             | : | :           |       | | |   :  |
Db     362 GGLKQVGEALRTGMVLALSIWDDYAANML--------WLDSNYPTNKDPSTPGVARGTCA 413

Qy     212 -----PAEIVARSGCKRNDDSSFPVFTPPSGGNGGT---------------GTPTSTAP 250
                ||:| |:|       ::: ||:    |:  |                 | ||::
Db     414 TTSGVPAQIEAQS------PNAYVVFSNIKFGDLNTTYTGTVSSSSVSSSHSSTSTSSSH 467

Qy     251 GSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQC 298
           |  | |   :| |  :| |||||::| ||| |  ||  || |||||
Db     468 SSSSTPPTQPTGVTVPQWGQCGGIGYTGSTTCASPYTCHVLNPYYSQC 515


RESULT 15
JU0150
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) I - Penicillium janthinellum
N;Alternate names: exo-cellobiohydrolase
C;Species: Penicillium janthinellum
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: JU0150; S21508
R;Christoph, A.K.; Weigel, T.O.; Schulz, G.
Gene 124, 57-65, 1993
A;Title: Cloning, sequencing, and heterologous expression of a cellulase-encoding cDN
A;Reference number: JU0150; MUID:93178976; PMID:8440481
A;Accession: JU0150
A;Molecule type: mRNA
A;Residues: 1-537 <CHR>
A;Cross-references: EMBL:X59054; NID:g3177; PIDN:CAA41780.1; PID:g3178
A;Note: the authors translated the codon ATG for residue 172 as Asn
C;Genetics:
A;Gene: cbh1
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain h
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;506-537/Domain: fungal cellulose-binding domain homology <FCB>

  Query Match             11.0%;  Score 183;  DB 1;  Length 537;
```

Best Local Similarity 23.4%; Pred. No. 1.1e-05;
Matches 79; Conservative 30; Mismatches 99; Indels 130; Gaps 14;

```
Qy      24 GSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSDFNVQSGCNGGSAYSCADQTP 83
           |:  : ||   | |                 || |  |:   |       |   :   ::|
Db     268 GTYSTDRYGGTCDPD-------------GCDFNPYRMGVTNFY-----GPGETIDTKSP 308

Qy      84 WAV------NDNLAYGFAA-------------------TSIAGGS-ESSWCCACYALTF 116
           : |      ||  : |  :                    ::| |   ||| |
Db     309 FTVVTQFLTNDGTSTGTLSEIKRFYVQGGKVIGNPQSTIVGVSGNSITDSWCNA------ 362

Qy     117 TSGPVAGKTMVVQSTSTGGDLGSNQFDI--AMPGGGVGIFNGC---------------- 157
                     |  ||  :|:|     | | | |: :|
Db     363 ------------QKSAFGD--TNEFSKHGGMAGMGAGLADGMVLVMSLWDDHASDMLWL 407

Qy     158 -----SSQFGGLPGAQYGGISSRDQCDSFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCP 212
                ::     |||:      |  ||    |             |    |:       |
Db     408 DSTYPTNATSTTPGAK------RGTCDISRRP------------NTVESTY-------P 441

Qy     213 AEIVARSGCKRNDDSSFPVFTPPSGGNGGTGTPTSTAPGSGQ----------TSPGGGS 261
              |  |  |    :|   ||  :  :  | | || :  :             |:  | |
Db     442 NAYVIYSNIKTGPLNS--TFTGGTTSSSSTTTTTSKSTSTSSSSKTTTTVTTTTTSSGSS 499

Qy     262 GCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           |  :: |||||| |::| |||||  || | ||:|||||
Db     500 GTGARDWAQCGGNGWTGPTTCVSPYTCTKQNDWYSQCL 537
```

Search completed: February 24, 2003, 11:52:59
Job time : 20 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 24, 2003, 11:43:08 ; Search time 14 Seconds
(without alignments)
885.816 Million cell updates/sec

Title: US-10-007-521-12
Perfect score: 1659
Sequence: 1 MRSTPVLRTTLAAALPLVAS.....TTCVSGTTCQKLNDYYSQCL 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	942.5	56.8	376	1	GUNK_FUSOX	P45699 fusarium ox
2	921.5	55.5	213	1	GUN5_HUMIN	P43316 humicola in
3	507	30.6	511	1	GUNB_PSEFL	P18126 pseudomonas
4	462	27.8	393	1	GUN1_USTMA	P54424 ustilago ma
5	221	13.3	242	1	GUN5_TRIRE	P43317 trichoderma
6	219.5	13.2	513	1	GUX1_TRIRE	P00725 trichoderma
7	215.5	13.0	513	1	GUX1_TRIVI	P19355 trichoderma
8	214	12.9	459	1	GUN1_TRIRE	P07981 trichoderma
9	203	12.2	463	1	GUN1_TRILO	Q12714 trichoderma
10	185	11.2	525	1	GUX1_HUMGR	P15828 humicola gr
11	184	11.1	516	1	GUX1_PHACH	P13860 phanerochae
12	183	11.0	537	1	GUX1_PENJA	Q06886 penicillium
13	181	10.9	516	1	GUX1_NEUCR	P38676 neurospora
14	172	10.4	320	1	CEL1_AGABI	Q00023 agaricus bi
15	159	9.6	344	1	GUN4_TRIRE	O14405 trichoderma
16	158.5	9.6	514	1	GUXC_FUSOX	P46238 fusarium ox
17	156	9.4	506	1	GUX2_AGABI	Q92400 agaricus bi
18	149.5	9.0	540	1	GUX1_ASPAC	O59843 aspergillus
19	143.5	8.6	388	1	GUN3_HUMIN	Q12624 humicola in
20	137	8.3	438	1	GUX3_AGABI	P49075 agaricus bi
21	129	7.8	418	1	GUN2_TRIRE	P07982 trichoderma
22	127	7.7	385	1	GUNF_FUSOX	P46239 fusarium ox
23	127	7.7	462	1	GUNB_FUSOX	P46236 fusarium ox
24	125.5	7.6	316	1	LORI_HUMAN	P23490 homo sapien
25	124	7.5	471	1	GUX2_TRIRE	P07987 trichoderma
26	122	7.4	182	1	KRUC_SHEEP	P26372 ovis aries
27	120.5	7.3	303	1	34KD_MYCTU	P71556 mycobacteri
28	120	7.2	5376	1	ZAN_MOUSE	O88799 mus musculu
29	118.5	7.1	210	1	PSBP_PORPU	P50272 porphyra pu
30	118.5	7.1	453	1	VG_DROME	Q26366 drosophila
31	118.5	7.1	1387	1	TROP_HUMAN	Q12816 homo sapien
32	115.5	7.0	491	1	YKC5_CAEEL	P41996 caenorhabdi
33	113.5	6.8	2090	1	N214_HUMAN	P35658 homo sapien
34	109.5	6.6	518	1	FUS_MOUSE	P56959 mus musculu
35	109	6.6	660	1	YHL1_EBV	P03181 epstein-bar
36	107	6.4	967	1	ATS1_RAT	Q9wuq1 rattus norv
37	107	6.4	968	1	ATS1_MOUSE	P97857 mus musculu
38	106.5	6.4	2704	1	G168_PARPR	P17053 paramecium
39	105.5	6.4	452	1	GUX1_CRYPA	Q00548 cryphonectr
40	105	6.3	481	1	LORI_MOUSE	P18165 mus musculu
41	103	6.2	491	1	YK98_MYCTU	Q10707 mycobacteri
42	103	6.2	537	1	SP70_DICDI	P15269 dictyosteli
43	103	6.2	677	1	SP87_DICDI	P54643 dictyosteli
44	103	6.2	1046	1	CHID_VIBFU	P96156 vibrio furn
45	102.5	6.2	914	1	WA22_MYCTU	O06794 mycobacteri

ALIGNMENTS

RESULT 1
GUNK_FUSOX
ID GUNK_FUSOX STANDARD; PRT; 376 AA.
AC P45699;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Putative endoglucanase type K precursor (EC 3.2.1.4) (Endo-1,4-beta-
DE glucanase) (Cellulase).
OS Fusarium oxysporum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; mitosporic Hypocreales; Fusarium.
OX NCBI_TaxID=5507;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95047531; PubMed=7959045;
RA Sheppard P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C.,
RA Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;
RT "The use of conserved cellulase family-specific sequences to clone
RT cellulase homologue cDNAs from Fusarium oxysporum.";
RL Gene 150:163-167(1994).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL
CC HYDROLASES).
CC --

CC --
DR EMBL; L29381; AAA65589.1; -.
DR HSSP; P43316; 2ENG.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR000334; GH_45.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 376 PUTATIVE ENDOGLUCANASE TYPE K.
FT DOMAIN 19 308 CATALYTIC.
FT DOMAIN 309 338 LINKER.
FT DOMAIN 339 376 CELLULOSE-BINDING.
FT ACT_SITE 29 29 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 140 140 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 376 AA; 39235 MW; B430A5F962B9F882 CRC64;

Query Match 56.8%; Score 942.5; DB 1; Length 376;
Best Local Similarity 49.5%; Pred. No. 2.9e-62;

```
  Matches  183;  Conservative   36;  Mismatches   72;  Indels   79;  Gaps    6;

Qy       9 TTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSDFNVQS 68
           | || | ||  ||||||| ||||||||||||:| |||||: |   || |   :|: |  :
Db       5 TLLALAGPLAVSAASGSGHSTRYWDCCKPSCSWSGKAAVNAPALTCDKNDNPISNTNAVN 64

Qy      69 GC-NGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMV 127
           ||  |||||:| : :|||||| |||||||| |:||||:||||||||||||:||| || |:
Db      65 GCEGGGSAYACTNYSPWAVNDELAYGFAATKISGGSEASWCCACYALTFTTGPVKGKKMI 124

Qy     128 VQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFG-GLPGAQYGGISSRDQCDSFPAPLK 186
           ||||:|||||| | ||: |||||||||:||:|:||  | |||||||||| :|||:|  ||
Db     125 VQSTNTGGDLGDNHFDLMMPGGGVGIFDGCTSEFGKALGGAQYGGISSRSECDSYPELLK 184

Qy     187 PGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVF-------------- 232
            || ||||||:||||| |||:|||||  ::  |||||:|||||| |
Db     185 DGCHWRFDWFENADNPDFTFEQVQCPKALLDISGCKRDDDSSFPAFKGDTSASKPQPSSS 244

Qy     233 --------------------------TPPS--------GGNGGTGTPTSTAPGSGQTSP 257
                                       | |:             |  | :| | : : :
Db     245 AKKTTSAAAAAQPQKTKDSAPVVQKSSTKPAAQPEPTKPADKPQTDKPVATKPAATKPAQ 304

Qy     258 GGGSGCTSQK------------------------WAQCGGIGF---SGCTTCVSGTTCQ 289
                 |:||                          : ||||      :|   | :|: |
Db     305 PVNKPKTTQKVRGTKTRGSCPAKTDATAKASVVPAYYQCGGSKSAYPNGNLACATGSKCV 364

Qy     290 KLNDYYSQCL 299
           | |:||||||:
Db     365 KQNEYYSQCV 374


RESULT 2
GUN5_HUMIN
ID   GUN5_HUMIN     STANDARD;      PRT;   213 AA.
AC   P43316;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Endoglucanase V (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DE   (Cellulase V) (EG V).
OS   Humicola insolens.
OC   Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX   NCBI_TaxID=34413;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rasmussen G., Mikkelsen J.-M., Schulein M., Patkar S.A., Hagen F.,
RA   Hjort C.M., Hastrup S.;
RT   "A cellulase preparation comprising an endoglucanase enzyme.";
RL   Patent number WO9117243, 14-NOV-1991.
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX   MEDLINE=93390621; PubMed=8377830;
RA   Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Dauter Z.,
RA   Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schuelein M.;
RT   "Structure and function of endoglucanase V.";
RL   Nature 365:362-364(1993).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX   MEDLINE=96101453; PubMed=8519779;
RA   Davies G.J., Tolley S.P., Henrissat B., Hjort C., Schuelein M.;
RT   "Structures of oligosaccharide-bound forms of the endoglucanase V
RT   from Humicola insolens at 1.9-A resolution.";
RL   Biochemistry 34:16210-16220(1995).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RA   Davies G.J., Dodson G.G., Moore M.H., Tolley S.P., Dauter Z.,
RA   Wilson K.S., Rasmussen G., Schuelein M.;
RT   "Structure determination and refinement of the Humicola insolens
RT   endoglucanase V at 1.5-A resolution.";
RL   Acta Crystallogr. D 52:7-17(1996).
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC       linkages in cellulose.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL
CC       HYDROLASES).
DR   PDB; 2ENG; 08-DEC-96.
DR   PDB; 3ENG; 16-JUN-97.
DR   PDB; 4ENG; 16-JUN-97.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.
FT   ACT_SITE     10     10       NUCLEOPHILE.
FT   ACT_SITE    121    121       PROTON DONOR.
SQ   SEQUENCE   213 AA;  22864 MW;  24334301BA3BC804 CRC64;

  Query Match             55.5%;  Score 921.5;  DB 1;  Length 213;
  Best Local Similarity   74.4%;  Pred. No. 5.8e-61;
  Matches  157;  Conservative   27;  Mismatches   26;  Indels    1;  Gaps    1;

Qy      26 GQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSDFNVQSGCN-GGSAYSCADQTPW 84
           |:|||||||||||| |  || |:|||::|:|||||::||: :|||  || ||||||||||
Db       3 GRSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQRITDFDAKSGCEPGGVAYSCADQTPW 62

Qy      85 AVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMVVQSTSTGGDLGSNQFDI 144
           ||||: | |||||||||| :|: |||||| |||||||||| |||||||||||||||| ||:
Db      63 AVNDDFALGFAATSIAGSNEAGWCCACYELTFTSGPVAGKKMVVQSTSTGGDLGSNHFDL 122

Qy     145 AMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCDSFPAPLKPGCQWRFDWFQNADNPTF 204
           :|||||||||:||: ||||||| :|||||||::|| ||  ||||| ||||||:|||||:|
Db     123 NIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECDRFPDALKPGCYWRFDWFKNADNPSF 182

Qy     205 TFQQVQCPAEIVARSGCKRNDDSSFPVFTPP 235
           :|:|||||||:|||:||:|||| :||    |
Db     183 SFRQVQCPAELVARTGCRRNDDGNFPAVQIP 213


RESULT 3
GUNB_PSEFL
ID   GUNB_PSEFL     STANDARD;      PRT;   511 AA.
AC   P18126;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE   (Cellulase) (EGB).
GN   CELB.
OS   Pseudomonas fluorescens.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=294;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 30-48.
RC   STRAIN=Sp. Cellulosa;
RX   MEDLINE=90355836; PubMed=2117693;
RA   Gilbert H.J., Hall J., Hazlewood G.P., Ferreira L.M.A.;
RT   "The N-terminal region of an endoglucanase from Pseudomonas
RT   fluorescens subspecies cellulosa constitutes a cellulose-binding
RT   domain that is distinct from the catalytic centre.";
RL   Mol. Microbiol. 4:759-767(1990).
CC   -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC       GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC       GLUCANS. EGB IS MOST ACTIVE AGAINST BARLEY BETA-GLUCAN, BUT SHOWED
CC       SIGNIFICANT ACTIVITY AGAINST AMORPHOUS AND CRISTALLINE CELLULOSE.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC       linkages in cellulose.
CC   -!- SUBCELLULAR LOCATION: Periplasmic.
CC   -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC       (CBD).
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL
CC       HYDROLASES).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X52615; CAA36844.1; -.
DR   PIR; S10527; S10527.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR001919; Bac_celose-bind.
DR   InterPro; IPR002883; CBD_5.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF00553; CBM_2; 1.
DR   Pfam; PF02013; CBM_10; 1.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS00561; CBD_BACTERIAL; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Signal; Periplasmic.
FT   SIGNAL       1     29
FT   CHAIN       30    511       ENDOGLUCANASE B.
FT   DOMAIN      30    131       CELLULOSE-BINDING (BY SIMILARITY).
FT   DOMAIN     132    173       SER-RICH (LINKER).
FT   DOMAIN     223    259       SER-RICH.
FT   DISULFID    32    127       BY SIMILARITY.
FT   ACT_SITE   276    276       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE   393    393       PROTON DONOR (BY SIMILARITY).
SQ   SEQUENCE   511 AA;  52078 MW;  3C3119D998291D8E CRC64;

  Query Match             30.6%;  Score 507;  DB 1;  Length 511;
  Best Local Similarity   39.6%;  Pred. No. 3.1e-30;
  Matches  107;  Conservative   38;  Mismatches   73;  Indels   52;  Gaps    7;

Qy       2 RSTPVLRTTLAAALPLVASAASG--------SGQSTRYWDCCKPSCAWPGKA-AVSQPVY 52
           ||:    ::|::|    ||: |          :| :|||||||| | |     ::  |:
Db     237 RSSVASSSSLSSATSSSASSVSSVPPIDGGCNGYATRYWDCCKPHCGWSANVPSLVSPLQ 296

Qy      53 ACDANFQRLSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACY 112
           :| ||  |||| :| | |:||  | | |: |:||:  ||||:|||     |   |  ||
Db     297 SCSANNTRLSDVSVGSSCDGGGGYMCWDKIPFAVSPTLAYGYAAT-----SSGDVCGRCY 351

Qy     113 ALTFT-----------SGPVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQF 161
             | ||           |  :||||:||:|: | |:   |||| :|||||| || ||:|:
Db     352 QLQFTGSSYNAPGDPGSAALAGKTMIVQATNIGYDVSGGQFDILVPGGGVGAFNACSAQW 411

Qy     162 G---GLPGAQYGGISSR------------------DQCDSF-----PAPLKPGCQWRFD 194
           |      ||||||  :                  ::|||         |: || |  :
Db     412 GVSNAELGAQYGGFLAACKQQLGYNASLSQYKSCVLNRCDSVFGSRGLTQLQQGCTWFAE 471

Qy     195 WFQNADNPTFTFQQVQCPAEIVARSGCKRN 224
           ||: ||||:  :::| ||||:  |||  |:
Db     472 WFEAADNPSLKYKEVPCPAELTTRSGMNRS 501


RESULT 4
GUN1_USTMA
ID   GUN1_USTMA     STANDARD;      PRT;   393 AA.
AC   P54424;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Endoglucanase 1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase 1)
DE   (Cellulase 1) (EG 1).
GN   EGL1.
OS   Ustilago maydis (Smut fungus).
OC   Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC   Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX   NCBI_TaxID=5270;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FBD11;
RX   MEDLINE=96145728; PubMed=8590631;
RA   Schauwecker F., Wanner G., Kahmann R.;
RT   "Filament-specific expression of a cellulase gene in the dimorphic
RT   fungus Ustilago maydis.";
RL   Biol. Chem. Hoppe-Seyler 376:617-625(1995).
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC       linkages in cellulose.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: HYPHAL TIP.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN FILAMENTOUS DIKARYON.
CC   -!- PTM: MAY ALSO BE O-GLYCOSYLATED.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL
CC       HYDROLASES).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; S81598; AAB36147.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Signal; Glycoprotein.
FT   SIGNAL       1     26       POTENTIAL.
FT   CHAIN       27    393       ENDOGLUCANASE 1.
FT   ACT_SITE    34     34       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE   152    152       PROTON DONOR (BY SIMILARITY).
FT   DOMAIN     270    385       ALA/GLY/SER-RICH.
FT   CARBOHYD   343    343       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   393 AA;  39594 MW;  65C753C610CD6AD3 CRC64;

  Query Match             27.8%;  Score 462;  DB 1;  Length 393;
  Best Local Similarity   37.7%;  Pred. No. 4.7e-27;
  Matches  117;  Conservative   41;  Mismatches  106;  Indels   46;  Gaps   16;

Qy      11 LAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRL----SDFNV 66
           |: :| ||      :| :|||||||  | :| ||| |  || || |:   |     | :
Db      12 LSLSLSLVHLDGVRAGMATRYWDCCLASASWEGKAPVYAPVDACKADGVTLIDSKKDPSG 71

Qy      67 QSGCNGGSAYSCADQTPW--AVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGK 124
           |||||||: : |:   |:    :  ||:|| |     | ||   |||:   |      ||
Db      72 QSGCNGGNKFMCSCMQPFDDETDPTLAFGFGA--FTTGQESDTDCACFYAEF-EHDAQGK 128

Qy     125 TM-----VVQSTSTGGDLGSNQFDIAMPGGGVGIF-NGCSSQFG---GLPGAQYGGISSR 175
           |      : | |: |||: |  ||  :||||:| |  || :|:|     | | ||||: |
Db     129 AMKRNKLIFQVTNVGGDVQSQNFDFQIPGGGLGAFPKGCPAQWGVEASLWGDQYGGVKSA 188

Qy     176 DQCDSFPAPLKPGCQWRF-DWFQNADNPTF--TFQQVQCPAEIVARSGCKRNDDSSFPVF 232
           :|   | ||: ||:||| :|    |||    : ::|:||  :: ||||:| ||::
Db     189 TECSKLPKPLQEGCKWRFSEW---GDNPVLKGSPKRVKCPKSLIDRSGCQRKDDNT---I 242

Qy     233 TPPSG--GNGGTGTPTS--------TAPG-SGQTSPGG--GSGCTSQKWAQCGGIGFSGC 279
           :| ||   :  |  |             | |  |::: ||  |||  |      ||   ||
Db     243 SPYSGKVDSANTAAPAQYKRDRSVCLAGGKKGKSAAGGVDGSGDAS------GGADASGA 296

Qy     280 TTCVSGTTCQ 289
               |:  |
Db     297 GGAAEGSQGQ 306


RESULT 5
GUN5_TRIRE
ID   GUN5_TRIRE     STANDARD;      PRT;   242 AA.
AC   P43317;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DE   (Cellulase V) (EG V).
GN   EGL5.
OS   Trichoderma reesei (Hypocrea jecorina).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Hypocreaceae; Hypocrea.
OX   NCBI_TaxID=51453;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=QM9414 / Rut C-30;
RX   MEDLINE=95075308; PubMed=7984103;
RA   Saloheimo A., Henrissat B., Hoffren A.-M., Teleman O., Penttilae M.;
RT   "A novel, small endoglucanase gene, egl5, from Trichoderma reesei
RT   isolated by expression in yeast.";
RL   Mol. Microbiol. 13:219-228(1994).
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC       linkages in cellulose.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z33381; CAA83846.1; -.
DR   HSSP; P00725; 2CBH.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF00734; CBM_1; 1.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   ProDom; PD001821; CBD_fungal; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; 1.
DR   PROSITE; PS50842; EXPANSIN_EG45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18    242       ENDOGLUCANASE V.
FT   ACT_SITE    27     27       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE   134    134       PROTON DONOR (BY SIMILARITY).
FT   DOMAIN      18    182       CATALYTIC.
FT   DOMAIN     183    205       PRO/SER-RICH (LINKER).
FT   DOMAIN     206    242       CELLULOSE-BINDING (BY SIMILARITY).
FT   CARBOHYD   182    182       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID   213    230       BY SIMILARITY.
FT   DISULFID   224    240       BY SIMILARITY.
SQ   SEQUENCE   242 AA;  24411 MW;  CC033FC51326C71D CRC64;

  Query Match             13.3%;  Score 221;  DB 1;  Length 242;
  Best Local Similarity   24.7%;  Pred. No. 1.3e-09;
  Matches   78;  Conservative   39;  Mismatches  101;  Indels   98;  Gaps   13;

Qy       7 LRTTLAAALPLVASAASGSGQSTRYWD-------CCKPSCAWPGKAAVSQPVYACDANFQ 59
           :: ||     :| : ::    :|||:|       |   | |:| :  :   ||   |  |
Db       1 MKATLVLGSLIVGAVSAYKATTTRYYDGQEGACGCGSSSGAFPWQLGIGNGVYTA-AGSQ 59

Qy      60 RLSDFNVQSGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFT-- 117
             | |    | |  |                                    |  || || |
Db      60 ALFDTAGASWCGAG--------------------------------CGKCYQLTSTGQ 85

Qy     118 -------SGPVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYG 170
                  :|  ||:::::|  |:   : |: |:   :  ||                 ||
Db      86 APCSSCGTGGAAGQSIIVMVTNLCPNNGNAQWCPVV--GGTN--------------QY- 127

Qy     171 GISSRDQCDSFPAPLKPGCQWRFD-WFQN---ADNPTFTFQQVQCPAEIVARSG---CKR 223
                       |  : ||   ||    ||    |: : || :  :  |   |
Db     128 ----------------GYSYHFDIMAQNEIFGDNVVVDFEPIACPGQAASDWGTCLCVG 170

Qy     224 NDDSSFPVFTPPSGGNGGTGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCV 283
           ::  |  ||  | : |:  | |: | :  : | ||   |   : |||| |::| |||
Db     171 QQETD-P--TPVLGNDTGSTPPGSSPPATSSSPPSGGGQQT--LYGQCGGAGWTGPTTCQ 225

Qy     284 SGTTCQKLNDYYSQCL 299
           :  ||:  | :|||||
Db     226 APGTCKVQNQWYSQCL 241
```

```
RESULT 6
GUX1_TRIRE
ID   GUX1_TRIRE     STANDARD;      PRT;   513 AA.
AC   P00725;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I)
DE   (CBHI) (1,4-beta-cellobiohydrolase).
GN   CBH1.
OS   Trichoderma reesei (Hypocrea jecorina), and
OS   Trichoderma koningii.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Hypocreaceae; Hypocrea.
OX   NCBI_TaxID=51453, 55202;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=T.reesei; STRAIN=L27;
RA   Shoemaker S., Schweickart V., Ladner M., Gelfand D., Kwok S.,
RA   Myambo K., Innis M.;
RT   "Molecular cloning of exo-cellobiohydrolase I derived from Trichoderma
RT   reesei strain L27.";
RL   Biotechnology 1:691-696(1983).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=T.koningii; STRAIN=G-39;
RX   MEDLINE=94100788; PubMed=7764306;
RA   Wey T.T., Hseu T.H., Huang L.;
RT   "Molecular cloning and sequence analysis of the cellobiohydrolase I
RT   gene from Trichoderma koningii G-39.";
RL   Curr. Microbiol. 28:31-39(1994).
RN   [3]
RP   ACTIVE SITE.
RC   SPECIES=T.reesei;
RA   Tomme P., Clayssens M.;
RT   "Identification of a functionally important carboxyl group in
RT   cellobiohydrolase I from Trichoderma reesei.";
RL   FEBS Lett. 243:239-243(1989).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.81 ANGSTROMS) OF 18-452.
RX   MEDLINE=94310436; PubMed=8036495;
RA   Divne C., Staahlberg J., Reinikainen T., Ruohonen L., Pettersson G.,
RA   Knowles J.K.C., Teeri T.T., Jones T.A.;
RT   "The three-dimensional crystal structure of the catalytic core of
RT   cellobiohydrolase I from Trichoderma reesei.";
RL   Science 265:524-528(1994).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 18-452.
RC   SPECIES=T.reesei; STRAIN=QM9414 / Rut C-30;
RX   MEDLINE=98128795; PubMed=9466911;
RA   Divne C., Staahlberg J., Teeri T.T., Jones T.A.;
RT   "High-resolution crystal structures reveal how a cellulose chain is
RT   bound in the 50 A long tunnel of cellobiohydrolase I from Trichoderma
RT   reesei.";
RL   J. Mol. Biol. 275:309-325(1998).
RN   [6]
RP   STRUCTURE BY NMR OF 478-513.
RC   SPECIES=T.reesei;
RX   MEDLINE=90057416; PubMed=2554967;
RA   Kraulis P.J., Clore G.M., Nilges M., Jones T.A., Pettersson G.,
RA   Knowles J., Gronenborn A.M.;
RT   "Determination of the three-dimensional solution structure of the C-
RT   terminal domain of cellobiohydrolase I from Trichoderma reesei. A
RT   study using nuclear magnetic resonance and hybrid distance
RT   geometry-dynamical simulated annealing.";
RL   Biochemistry 28:7241-7257(1989).
RN   [7]
RP   STRUCTURE BY NMR OF 478-513.
RC   SPECIES=T.reesei;
RX   MEDLINE=97194052; PubMed=9041630;
RA   Mattinen M.L., Kontteli M., Kerovuo J., Linder M., Annila A.,
```

```
RA   Lindeberg G., Reinikainen T., Drakenberg T.;
RT   "Three-dimensional structures of three engineered cellulose-binding
RT   domains of cellobiohydrolase I from Trichoderma reesei.";
RL   Protein Sci. 6:294-303(1997).
CC   -!- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC       GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC       (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC       (2) EXOCELLOBIOHYDROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
CC          FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC       (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC          SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC       in cellulose and cellotetraose, releasing cellobiose from the non-
CC       reducing ends of the chains.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: T.REESEI PRODUCES TWO DIFFERENT
CC       EXOCELLOBIOHYDROLASES. THEY ARE UNIQUE IN THAT THEY CAN HYDROLYSE
CC       CRYSTALLINE CELLULOSE IN THE ABSENCE OF ENDOGLUCANASES.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69976; CAA49596.1; -.
DR   PIR; A00902; EUTQI.
DR   PIR; S45380; S45380.
DR   PDB; 1CEL; 01-NOV-94.
DR   PDB; 2CEL; 12-MAR-97.
DR   PDB; 3CEL; 12-MAR-97.
DR   PDB; 4CEL; 12-MAR-97.
DR   PDB; 5CEL; 24-DEC-97.
DR   PDB; 6CEL; 24-DEC-97.
DR   PDB; 7CEL; 24-DEC-97.
DR   PDB; 8CEL; 25-FEB-98.
DR   PDB; 1CBH; 15-JAN-90.
DR   PDB; 2CBH; 15-JAN-90.
DR   PDB; 1AZ6; 08-APR-98.
DR   PDB; 1AZH; 08-APR-98.
DR   PDB; 1AZJ; 29-APR-98.
DR   PDB; 1AZK; 29-APR-98.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR001722; GH_7.
DR   Pfam; PF00734; CBM_1; 1.
DR   Pfam; PF00840; Glyco_hydro_7; 1.
DR   ProDom; PD001821; CBD_fungal; 1.
DR   ProDom; PD186135; GH_7; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal;
KW   3D-structure.
FT   SIGNAL       1     17
FT   CHAIN       18    513       EXOGLUCANASE I.
FT   DOMAIN      18    453       CATALYTIC.
FT   DOMAIN     454    477       LINKER.
FT   DOMAIN     478    513       CELLULOSE-BINDING.
FT   ACT_SITE   143    143       PROBABLE.
FT   ACT_SITE   229    229       NUCLEOPHILE.
FT   ACT_SITE   234    234       PROTON DONOR.
FT   MOD_RES     18     18       PYRROLIDONE CARBOXYLIC ACID.
FT   CARBOHYD    62     62       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    81     81       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   287    287       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   401    401       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    21     89
FT   DISULFID    36     42
FT   DISULFID    67     88
FT   DISULFID    78     84
FT   DISULFID   155    414
FT   DISULFID   189    227
FT   DISULFID   193    226
FT   DISULFID   247    273
FT   DISULFID   255    260
FT   DISULFID   278    348
FT   DISULFID   485    502
FT   DISULFID   496    512
FT   STRAND     479    479
FT   TURN       481    482
FT   STRAND     484    486
FT   TURN       488    489
FT   STRAND     496    496
FT   TURN       498    499
FT   STRAND     502    506
FT   TURN       507    508
FT   STRAND     509    512
SQ   SEQUENCE   513 AA;  54073 MW;  9F5C0A8A854F2C12 CRC64;

  Query Match             13.2%;  Score 219.5;  DB 1;  Length 513;
  Best Local Similarity   26.9%;  Pred. No. 3.5e-09;
  Matches   91;  Conservative   36;  Mismatches  122;  Indels   89;  Gaps   22;

Qy      16 PLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPV--YACDANFQRLSDFNVQSGCNGG 73
           |   :| :| |    :  ||     |    ::|: :  : |    | : :    || |
Db     211 PSSNNANTGIG---GHGSCCSEMDIWEAN-SISEALTPHPCTTVGQEICE---GDGC--G 261

Qy      74 SAYS-------C----ADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVA 122
             ||       |     |  |: : :   ||       | |          : |  |
Db     262 GTYSDNRYGGTCDPDGCDWNPYRLGNTSFYG------PGSS--------FTLDTTK---- 303

Qy     123 GKTMVVQSTSTGGDLG----SNQFDIAMPGGGVGIFNG-------CS---SQFGGLPGAQ 168
            |  ||    | | :      |      |   :| ::|       |:   ::|||   :
Db     304 -KLTVVTQFETSGAINRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSSFSD 362

Qy     169 YGGISSRDQCDSFPAPLKPGCQWRFDWFQN---ADNPTFTFQQVQCPAEIVARSGCKRND 225
           ||::   :  |    |     |  |:: |    |:   | :    |  :  |  |  :
Db     363 KGGLTQFKKATSGGMVLVMSL-WD-DYYANMLWLDSTYPTNETSSTPGAV--RGSCSTSS 418

Qy     226 ------DSSFP---------VFTP------PSGGN---GGTGTPTSTAPGSGQTSPGGGS 261
                :|  |           | |       |||||    |  || |:  |     |: |
Db     419 GVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSGGNPPGGNRGTTTTRRPA---TTTGSSP 475

Qy     262 GCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           | |   : ||||||:|| | | |||||| || ||||||
Db     476 GPTQSHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL 513


RESULT 7
GUX1_TRIVI
ID   GUX1_TRIVI     STANDARD;      PRT;   513 AA.
AC   P19355;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase) (1,4-
DE   beta-cellobiohydrolase).
GN   CBH1.
OS   Trichoderma viride.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; mitosporic Hypocreales; Trichoderma.
OX   NCBI_TaxID=5547;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91016856; PubMed=2216737;
RA   Cheng C., Tsukagoshi N., Udaka S.;
RT   "Nucleotide sequence of the cellobiohydrolase gene from Trichoderma
RT   viride.";
RL   Nucleic Acids Res. 18:5559-5559(1990).
CC   -!- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC       GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC       (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
```

```
CC       (2) EXOCELLOBIOHYDROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
CC           FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC       (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC           SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC       in cellulose and cellotetraose, releasing cellobiose from the non-
CC       reducing ends of the chains.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53931; CAA37878.1; -.
DR   PIR; S11439; S11439.
DR   HSSP; P00725; 8CEL.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR001722; GH_7.
DR   Pfam; PF00734; CBM_1; 1.
DR   Pfam; PF00840; Glyco_hydro_7; 1.
DR   ProDom; PD001821; CBD_fungal; 1.
DR   ProDom; PD186135; GH_7; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT   SIGNAL       1     17
FT   CHAIN       18    513       EXOGLUCANASE I.
FT   DOMAIN      18    453       CATALYTIC.
FT   DOMAIN     454    477       LINKER.
FT   DOMAIN     478    513       CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE   229    229       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE   234    234       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    62     62       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    81     81       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   287    287       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID   485    502       BY SIMILARITY.
FT   DISULFID   496    512       BY SIMILARITY.
SQ   SEQUENCE   513 AA;  53891 MW;  58EF5552717C4C8E CRC64;

  Query Match             13.0%;  Score 215.5;  DB 1;  Length 513;
  Best Local Similarity   26.2%;  Pred. No. 6.8e-09;
  Matches   86;  Conservative   36;  Mismatches  137;  Indels   69;  Gaps   16;

Qy      16 PLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPV--YACDANFQRLSDFNVQSGCNGG 73
           |   :| :| |    :  ||     |    ::|: :  : |    | : : :   |   |
Db     211 PSSNNANTGIG---GHGSCCSEMDIWEAN-SISEALTPHPCTTVGQEICEGDSCGGTYSG 266

Qy      74 SAY--SC----ADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMV 127
             |  :|     |  |: : :   ||       | |        : |  |      |  |
Db     267 DRYGGTCDPDGCDWNPYRLGNTSFYG------PGSS--------FTLDTTK-----KLTV 307

Qy     128 VQSTSTGGDLG----SNQFDIAMPGGGVGIFNGCS----------SQFGGLPGAQYGGIS 173
           |    | | :      |      |   :| ::| |          ::|||   :  ||::
Db     308 VTQFETSGAINRYYVQNGVTFQQPNAELGDYSGNSLDDDYCAAEEAEFGGSSFSDKGGLT 367

Qy     174 SRDQCDSFPAPLKPGCQWRFDWFQN---ADNPTFTFQQVQCPAEI----VARSGCKRNDD 226
             :  |    |     |  |:: |    |:   | :    |  :       ||     :
Db     368 QFKKATSGGMVLVMSL-WD-DYYANMLWLDSTYPTDETSSTPGAVRGSSSTSSGVPAQLE 425

Qy     227 SSFP---------VFTP------PSGGNGGTGTPTSTAPGSGQTSPGGGSGCTSQKWAQC 271
           |: |          | |      |||||   | |  |      || |   | |   : ||
Db     426 SNSPNAKVVYSNIKFGPIGSTGNPSGGNPPGGNPPGTTTPRPATSTGSSPGPTQTHYGQC 485

Qy     272 GGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           ||||: | | | ||:||| || ||||||
Db     486 GGIGYIGPTVCASGSTCQVLNPYYSQCL 513
```

```
RESULT 8
GUN1_TRIRE
ID   GUN1_TRIRE     STANDARD;      PRT;   459 AA.
AC   P07981;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Endoglucanase EG-1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE   (Cellulase).
GN   EGL1.
OS   Trichoderma reesei (Hypocrea jecorina).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Hypocreaceae; Hypocrea.
OX   NCBI_TaxID=51453;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VTT-D-80133;
RX   MEDLINE=87106822; PubMed=2948877;
RA   Penttilae M., Lehtovaara P., Nevalainen H., Bhikhabhai R.,
RA   Knowles J.K.C.;
RT   "Homology between cellulase genes of Trichoderma reesei: complete
RT   nucleotide sequence of the endoglucanase I gene.";
RL   Gene 45:253-263(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=L27;
RA   van Arsdell J.N., Kwok S., Schweickart V.L., Ladner M.B.,
RA   Gelfand D.H., Innis M.A.;
RT   "Cloning, characterization, and expression in Saccharomyces cerevisiae
RT   of endoglucanase I from Trichoderma reesei.";
RL   Biotechnology 5:60-64(1987).
RN   [3]
RP   ACTIVE SITE GLU-149.
RA   Tomme P., Clayssens M.;
RT   "Identification of a functionally important carboxyl group in
RT   cellobiohydrolase I from Trichoderma reesei.";
RL   FEBS Lett. 243:239-243(1989).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (3.6 ANGSTROMS) OF 23-393.
RX   MEDLINE=97467423; PubMed=9325098;
RA   Kleywegt G.J., Zou J.-Y., Divne C., Davies G.J., Sinning I.,
RA   Staehlberg J., Reinikainen T., Srisodsuk M., Teeri T.T., Jones T.A.;
RT   "The crystal structure of the catalytic core domain of endoglucanase
RT   I from Trichoderma reesei at 3.6-A resolution, and a comparison with
RT   related enzymes.";
RL   J. Mol. Biol. 272:383-397(1997).
CC   -!- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC       GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC       (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC       (2) EXOCELLOBIOHYDROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
CC           FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC       (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC           SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC       linkages in cellulose.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M15665; AAA34212.1; -.
DR   PIR; A25928; A25928.
DR   PIR; A25565; A25565.
DR   PDB; 1EG1; 19-MAR-99.
```

```
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR001722; GH_7.
DR   Pfam; PF00734; CBM_1; 1.
DR   Pfam; PF00840; Glyco_hydro_7; 1.
DR   ProDom; PD186135; GH_7; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal;
KW   3D-structure.
FT   SIGNAL        1     22
FT   CHAIN        23    459       ENDOGLUCANASE EG-1.
FT   DOMAIN       23    397       CATALYTIC.
FT   DOMAIN      398    423       LINKER.
FT   DOMAIN      424    459       CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE    149    149
FT   ACT_SITE    218    218       NUCLEOPHILE.
FT   ACT_SITE    223    223       PROTON DONOR.
FT   CARBOHYD     78     78       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    164    164       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    204    204       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    208    208       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    394    394       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    431    448       BY SIMILARITY.
FT   DISULFID    442    458       BY SIMILARITY.
SQ   SEQUENCE   459 AA;  48208 MW;  D235A256F808CBB9 CRC64;

  Query Match             12.9%;  Score 214;  DB 1;  Length 459;
  Best Local Similarity   27.0%;  Pred. No. 7.9e-09;
  Matches  108;  Conservative   31;  Mismatches  101;  Indels  160;  Gaps   27;

Qy      19 ASAASGSSGQS---TRYWDCCKPSCAWPGKAAVSQPVYACDA---------NFQRLSDFNV 66
           ||  : || |    :|     || :  | ::||  :|  |:          | | || |:|
Db     101 ASGVTTSGSSLTMNQY----MPSSS-GGYSSVSPRLYLLDSDGEYVMLKLNGQELS-FDV 154

Qy      67 Q--------------------------SGCNGGSAYSCADQTP---WAVNDNLAYGFA 95
                                        :| | || | |  | |   |  |  |
Db     155 DLSALPCGENGSLYLSQMDENGGANQYNTAGANYGSGY-CDAQCPVQTWR-NGTL----- 207

Qy      96 ATSIAGGSESSWCC----------------------AC---------YALTFTS--GPV 121
                |   :||                        ||         |   : |  ||
Db     208 -----NTSHQGFCCNEMDILEGNSRANALTPHSCTATACDSAGCGFNPYGSGYKSYYGP- 261

Qy     122 AGKTMVVQST-----------STGGDLGS-------NQFDI--AMPGGGVGIFNGCSSQ 160
             | |:   |           |  |:| |       |  ||  | |||     : | |
Db     262 -GDTVDTSKTFTIITQFNTDNGSPSGNLVSITRKYQQNGVDIPSAQPGGDT--ISSCPS- 317

Qy     161 FGGLPGAQYGGISSRDQCDSFPAPLKPGCQWRFD-WFQNAD--NPTFTFQQVQC------ 211
                : |||:::  :       |  |    |  |  |:  |   :     |
Db     318 -----ASAYGGLATMGKA------LSSGMVLVFSIWNDNSQYMNWLDSGNAGPCSSTEGN 366

Qy     212 PAEIVARSGCKRNDDSSFPVFTPPSGGNGGTGTPTSTA----PGSGQT--------SPGG 259
           |: |:|      |: ::  ||:    |: |: |  |||    | |  |          :
Db     367 PSNILA------NNPNTHVVFSNIRWGDIGS-TTNSTAPPPPPASSTTFSTTRRSSTTSS 419

Qy     260 GSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
             ||   | ||||||:||| || ||||||  ||||||||
Db     420 SPSCTQTHWGQCGGIGYSGCKTCTSGTTCQYSNDYYSQCL 459


RESULT 9
GUN1_TRILO
ID   GUN1_TRILO     STANDARD;      PRT;   463 AA.
AC   Q12714;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Endoglucanase EG-1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE   (Cellulase).
GN   EGL1.
OS   Trichoderma longibrachiatum.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; mitosporic Hypocreales; Trichoderma.
OX   NCBI_TaxID=5548;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CECT 2606;
RA   Perez-Gonzalez J.A.;
RL   Submitted (NOV-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC       GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC       (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC       (2) EXOCELLOBIOHYDROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
CC          FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC       (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC          SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC       linkages in cellulose.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X60652; CAA43059.1; -.
DR   HSSP; P07981; 1EG1.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR001722; GH_7.
DR   Pfam; PF00734; CBM_1; 1.
DR   Pfam; PF00840; Glyco_hydro_7; 1.
DR   ProDom; PD186135; GH_7; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23    463       ENDOGLUCANASE EG-1.
FT   DOMAIN       23    397       CATALYTIC.
FT   DOMAIN      402    427       LINKER.
FT   DOMAIN      428    463       CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE    218    218       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE    223    223       PROTON DONOR (BY SIMILARITY).
FT   DISULFID    435    452       BY SIMILARITY.
FT   DISULFID    446    462       BY SIMILARITY.
FT   CARBOHYD     78     78       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    164    164       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    204    204       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    208    208       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    394    394       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   463 AA;  48337 MW;  B3AC3DFD3ADD2B1C CRC64;

  Query Match             12.2%;  Score 203;  DB 1;  Length 463;
  Best Local Similarity   28.4%;  Pred. No. 5.1e-08;
  Matches   86;  Conservative   22;  Mismatches  101;  Indels   94;  Gaps   18;

Qy      68 SGCNGGSAYSCADQTP---WAVNDNL---AYGFAAT--SIAGGSESSWCCACYALTFTSG 119
           :| | || | |  | |   |  |  |     ||      |  |:  :     :: | |:
Db     184 AGANYGSGY-CDAQCPVQTWR-NGTLNTSGQGFCCNEMDILEGNSRANALTPHSCTATAC 241

Qy     120 PVAG----------------------KTMVV------QSTSTGGDLGS-------NQFD 143
             ||                      ||  :      : |  |:| |       |  |
Db     242 DSAGCGFNPYGSGYPNYFGPGDTVDTSKTFTIITQFNTDNGSPSGNLVSITRKYRQNGVD 301

Qy     144 I--AMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCDSFPAPLKPGCQWRFD-WFQNAD 200
           |  | |||     : | |       : |||::  :       |  |    |  |  |:
Db     302 IPSAKPGGDT--ISSCPS------ASAYGGLATMGKA------LSSGMVLVFSIWNDNSQ 347

Qy     201 -----------------NPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGN--- 239
                            ||:       |  | |  :  | |    |  :|||
Db     348 YMNWLDSGRAGPCSSTEGNPSNIL--ANNPGTHVVYSNIRWGDIGS---TTNSTGGNPPP 402
```

```
Qy     240 ---GGTGTPTSTAPGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYS 296
           :  |  ||    |  ||       ||   | ||||||::|| || ||||||  |||||
Db     403 PPPPASSTTFSTTRRSSTTS--SSPSCTQTHWGQCGGIGYTGCKTCTSGTTCQYGNDYYS 460

Qy     297 QCL 299
           |||
Db     461 QCL 463


RESULT 10
GUX1_HUMGR
ID   GUX1_HUMGR     STANDARD;      PRT;   525 AA.
AC   P15828;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I) (1,4-
DE   beta-cellobiohydrolase) (Beta-glucancellobiohydrolase).
GN   CBH-1.
OS   Humicola grisea.
OC   Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX   NCBI_TaxID=5527;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Var. thermoidea;
RX   MEDLINE=90175006; PubMed=2308855;
RA   de Oliviera Alzevedo M., Radford A.;
RT   "Sequence of cbh-1 gene of Humicola grisea var. thermoidea.";
RL   Nucleic Acids Res. 18:668-668(1990).
CC   -!- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC       GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC       (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC       (2) EXOCELLOBIOHYDROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
CC           FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC       (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC           SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC       in cellulose and cellotetraose, releasing cellobiose from the non-
CC       reducing ends of the chains.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X17258; CAA35159.1; -.
DR   PIR; S08240; S08240.
DR   HSSP; P00725; 8CEL.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR001722; GH_7.
DR   Pfam; PF00734; CBM_1; 1.
DR   Pfam; PF00840; Glyco_hydro_7; 1.
DR   ProDom; PD001821; CBD_fungal; 1.
DR   ProDom; PD186135; GH_7; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    525       EXOGLUCANASE I.
FT   DOMAIN       19    467       CATALYTIC.
FT   DOMAIN      468    489       LINKER.
FT   DOMAIN      490    525       CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE    231    231       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE    236    236       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    289    289       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    497    514       BY SIMILARITY.
FT   DISULFID    508    524       BY SIMILARITY.

SQ   SEQUENCE   525 AA;  55693 MW;  A6684D4CF881E090 CRC64;

  Query Match             11.2%;  Score 185;  DB 1;  Length 525;
  Best Local Similarity   24.9%;  Pred. No. 1.2e-06;
  Matches   97;  Conservative   33;  Mismatches  136;  Indels  124;  Gaps   23;

Qy      11 LAAALPLVASAASGSGQSTRYWDCCKPSCAWPG-KAAVSQPVYACDANFQRLSDF----- 64
           |  ||  |:  | |    :||          || ||        |||   |   |
Db     159 LNGALYFVSMDADGG--LSRY----------PGNKAGAKYGTGYCDAQCPRDIKFINGEA 206

Qy      65 NVQSGCN-------GGSAYS--CADQTPWAVNDNLAYGFA--ATSIAGGS--ESSWCCAC 111
           |::           |   |   |::   |  | |:|  |      :| | |  |   |
Db     207 NIEGWTGSTNDPNAGAGRYGTCCSEMDIWEAN-NMATAFTPHPCTIIGQSRCEGDSCGGT 265

Qy     112 YALTFTSG-----------------PVAGKTMVVQST-----------STGGDLGS-NQF 142
           |:   :|                   || | | :|             ||||   :|
Db     266 YSNERYAGVCDPDGCDFNSYRQGNKTFYGKGMTVDTTKKITVVTQFLKDANGDLGEIKRF 325

Qy     143 -----------DIAMPG--GGVGIFNGCSSQ---FGGLPG-AQYGGISSRDQCDSFPAPL 185
                      :  :||  |     : |  |   || :    : ||:    :  : |\ |
Db     326 YVQDGKIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMGKALAGPMVL 385

Qy     186 KPGCQWRFDWFQNADNPTF---TFQQVQCPAEIVARSGCKRN---DDSSFPV-------- 231
                  |  :| |  :   ||     | :   : | :|      |  |
Db     386 VMSI-----WDDHASNMLWLDSTF-----PVDAAGKPGAERGACPTTSGVPAEVEAEAPN 435

Qy     232 ---------FTP--------PSGGNGGTG-----TPTSTAPGSGQTSPGGGSGCTSQKWA 269
                   | |         |  ||||        ||:|   :  |:    :|  : :|
Db     436 SNVVFSNIRFGPIGSTVAGLPGAGNGGNNGGNPPPPTTTTSSAPATTTTASAGPKAGRWQ 495

Qy     270 QCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           |||||||:| | |     | ||||:|||||
Db     496 QCGGIGFTGPTQCEEPYICTKLNDWYSQCL 525


RESULT 11
GUX1_PHACH
ID   GUX1_PHACH     STANDARD;      PRT;   516 AA.
AC   P13860;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I)
DE   (1,4-beta-cellobiohydrolase).
GN   CBH1.
OS   Phanerochaete chrysosporium.
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC   Aphyllophorales; Corticiaceae; Phanerochaete.
OX   NCBI_TaxID=5306;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ME446;
RX   MEDLINE=89232732; PubMed=3246351;
RA   Sims P.F.G., James C., Broda P.;
RT   "The identification, molecular cloning and characterisation of a gene
RT   from Phanerochaete chrysosporium that shows strong homology to the
RT   exo-cellobiohydrolase I gene from Trichoderma reesei.";
RL   Gene 74:411-422(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ME446;
RX   MEDLINE=94335641; PubMed=8057846;
RA   Sims P.F.G., Soares-Felipe M.S., Wang Q., Gent M.E., Tempelaars C.,
RA   Broda P.;
RT   "Differential expression of multiple exo-cellobiohydrolase I-like
RT   genes in the lignin-degrading fungus Phanerochaete chrysosporium.";
RL   Mol. Microbiol. 12:209-216(1994).
CC   -!- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC       GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC       (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC       (2) EXOCELLOBIOHYDROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
CC           FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC       (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
```

```
CC          SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC       in cellulose and cellotetraose, releasing cellobiose from the non-
CC       reducing ends of the chains.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M22220; AAB46373.1; -.
DR   EMBL; Z22528; CAA80253.1; -.
DR   PIR; JS0083; JS0083.
DR   HSSP; P00725; 2CBH.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR001722; GH_7.
DR   Pfam; PF00734; CBM_1; 1.
DR   Pfam; PF00840; Glyco_hydro_7; 1.
DR   ProDom; PD001821; CBD_fungal; 1.
DR   ProDom; PD186135; GH_7; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    516       EXOGLUCANASE I.
FT   DOMAIN        ?    449       CATALYTIC.
FT   DOMAIN      450    480       LINKER.
FT   DOMAIN      481    516       CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE    225    225       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE    230    230       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    208    208       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    326    326       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    442    442       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    488    505       BY SIMILARITY.
FT   DISULFID    499    515       BY SIMILARITY.
FT   CONFLICT     27     28       RS -> EN (IN REF. 1).
FT   CONFLICT     30     31       PA -> RT (IN REF. 1).
SQ   SEQUENCE   516 AA;  54857 MW;  1C7C3D338ECE1B72 CRC64;

  Query Match             11.1%;  Score 184;  DB 1;  Length 516;
  Best Local Similarity   24.7%;  Pred. No. 1.4e-06;
  Matches   86;  Conservative   37;  Mismatches  113;  Indels  112;  Gaps   20;

Qy      20 SAASGSGQSTRYWDCCKPSCAWPGK--AAVSQPVYACDANFQRLSDFNVQSGCNGG---- 73
           || :|:|    |  ||     |     ||   | : |          | |: |:|
Db     211 SANAGTG---NYGTCCTEMDIWEANNDAAAYTP-HPCTT--------NAQTRCSGSDCTR 258

Qy      74 ------------SAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPV 121
                      :::   |||           |    :   :   :      :|   |
Db     259 DTGLCDADGCDFNSFRMGDQT----------FLGKGLTVDTSKPFTVVTQFIT-NDGTS 306

Qy     122 AG-----KTMVVQSTSTGGDLGSNQFDIAMPG-GGVGIF--NGCSSQ---FGGLP-GAQY 169
           ||     : : ||:    | :  |   : :||   |     | || |   ||     ||:
Db     307 AGTLTEIRRLYVQN----GKVIQNS-SVKIPGIDPVNSITDNFCSQQKTAFGDTNYFAQH 361

Qy     170 GGISSRDQC------------DSFPAPLKPGCQWRFDWFQ-----NADNPTFTFQQVQC- 211
           ||:    :             | : | :         |       | |  |    :  |
Db     362 GGLKQVGEALRTGMVLALSIWDDYAANML--------WLDSNYPTNKDPSTPGVARGTCA 413

Qy     212 -----PAEIVARSGCKRNDDSSFPVFTPPSGGNGGT---------------GTPTSTAP 250
                ||:| |:|       ::: ||:    |:  |                 | ||::
Db     414 TTSGVPAQIEAQS------PNAYVVFSNIKFGDLNTTYTGTVSSSSVSSSHSSTSTSSSH 467

Qy     251 GSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQC 298
           |  | |   :| |  :| |||||::| ||| |  ||  || |||||
Db     468 SSSSTPPTQPTGVTVPQWGQCGGIGYTGSTTCASPYTCHVLNPYYSQC 515
```

```
RESULT 12
GUX1_PENJA
ID   GUX1_PENJA     STANDARD;      PRT;   537 AA.
AC   Q06886;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I)
DE   (1,4-beta-cellobiohydrolase).
GN   CBH1.
OS   Penicillium janthinellum (Penicillium vitale).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX   NCBI_TaxID=5079;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C41;
RX   MEDLINE=93178976; PubMed=8440481;
RA   Koch A., Weigel C.T.O., Schulz G.;
RT   "Cloning, sequencing, and heterologous expression of a cellulase-
RT   encoding cDNA (cbh1) from Penicillium janthinellum.";
RL   Gene 124:57-65(1993).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC       in cellulose and cellotetraose, releasing cellobiose from the non-
CC       reducing ends of the chains.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59054; CAA41780.1; -.
DR   PIR; JU0150; JU0150.
DR   HSSP; P00725; 8CEL.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR001722; GH_7.
DR   Pfam; PF00734; CBM_1; 1.
DR   Pfam; PF00840; Glyco_hydro_7; 1.
DR   ProDom; PD001821; CBD_fungal; 1.
DR   ProDom; PD186135; GH_7; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    537       EXOGLUCANASE I.
FT   DOMAIN       19    453       CATALYTIC.
FT   DOMAIN      454    477       LINKER.
FT   DOMAIN      478    537       CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE    235    235       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE    240    240       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    136    136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    456    456       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    509    526       BY SIMILARITY.
FT   DISULFID    520    536       BY SIMILARITY.
SQ   SEQUENCE   537 AA;  56842 MW;  A6B9C6EB73F17FE4 CRC64;

  Query Match             11.0%;  Score 183;  DB 1;  Length 537;
  Best Local Similarity   23.4%;  Pred. No. 1.7e-06;
  Matches   79;  Conservative   30;  Mismatches   99;  Indels  130;  Gaps   14;

Qy      24 GSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSDFNVQSGCNGGSAYSCADQTP 83
             |:  : ||   | |                 || |  |:   |       |   :   ::|
Db     268 GTYSTDRYGGTCDPD-------------GCDFNPYRMGVTNFY-----GPGETIDTKSP 308
```

```
Qy      84 WAV------NDNLAYGFAA--------------------TSIAGGS-ESSWCCACYALTF 116
           : |      ||  : |  :                    ::| |   ||| |
Db     309 FTVVTQFLTNDGTSTGTLSEIKRFYVQGGKVIGNPQSTIVGVSGNSITDSWCNA------ 362

Qy     117 TSGPVAGKTMVVQSTSTGGDLGSNQFDI--AMPGGGVGIFNGC----------------- 157
                       |  ||  :|:|     | | | |: :|
Db     363 ------------QKSAFGD--TNEFSKHGGMAGMGAGLADGMVLVMSLWDDHASDMLWL 407

Qy     158 -----SSQFGGLPGAQYGGISSRDQCDSFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCP 212
                ::     |||:      |  ||    |              |    |:       |
Db     408 DSTYPTNATSTTPGAK------RGTCDISRRP------------NTVESTY-------P 441

Qy     213 AEIVARSGCKRNDDSSFPVFTPPSGGNGGTGTPTSTAPGSGQ-----------TSPGGGS 261
           |  |  |    :|   ||  :  :  | | || :  :              |:  | |
Db     442 NAYVIYSNIKTGPLNS--TFTGGTTSSSSTTTTSKSTSTSSSSKTTTTVTTTTTSSGSS 499

Qy     262 GCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           |  :: |||||| |::| |||||  || | ||:|||||
Db     500 GTGARDWAQCGGNGWTGPTTCVSPYTCTKQNDWYSQCL 537
```

RESULT 13
GUX1_NEUCR
ID GUX1_NEUCR STANDARD; PRT; 516 AA.
AC P38676;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Exoglucanase 1 precursor (EC 3.2.1.91) (Exocellobiohydrolase 1)
DE (1,4-beta-cellobiohydrolase).
GN CBH-1.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=95369725; PubMed=7642129;
RA Taleb F., Radford A.;
RT "The cellulase complex of Neurospora crassa: cbh-1 cloning,
RT sequencing and homologies.";
RL Gene 161:137-138(1995).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
CC HYDROLASES).
CC -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X77778; CAA54815.1; -.
DR PIR; S42093; S42093.
DR HSSP; P00725; 8CEL.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR001722; GH_7.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF00840; Glyco_hydro_7; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD186135; GH_7; 1.
DR SMART; SM00236; fCBD; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 516 EXOGLUCANASE 1.
FT DOMAIN 18 445 CATALYTIC.
FT DOMAIN 446 480 LINKER.
FT DOMAIN 481 516 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 223 223 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 228 228 PROTON DONOR (BY SIMILARITY).
FT DISULFID 488 505 BY SIMILARITY.
FT DISULFID 499 515 BY SIMILARITY.
SQ SEQUENCE 516 AA; 54471 MW; 38E598406EA81900 CRC64;

Query Match 10.9%; Score 181; DB 1; Length 516;
Best Local Similarity 24.5%; Pred. No. 2.3e-06;
Matches 81; Conservative 43; Mismatches 129; Indels 78; Gaps 16;

```
Qy      22 ASGSGQSTRYWDCCKPSCAW-PGKAAVSQPVYACDANFQRLSDFNVQSGCNGGSAYS--- 77
           |:| |    :  ||      |   | : :   : |    | : :     | : |  ||
Db     211 ANGIGD---HGSCCSEMDIWEANKVSTAFTPHPCTTIEQHMCE-----GDSCGGTYSDDR 262

Qy      78 ----C----ADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAG-KTMVV 128
                |    |   : : :   ||   |     : | :       :  ::| :|  |   |
Db     263 YGVLCDADGCDFNSYRMGNTTFYGEGKTV---DTSSKFTVVTQFIKDSAGDLAEIKAFYV 319

Qy     129 QSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQ---FGGLPG-AQYGGISSRDQC------ 178
           |:   |  : ::| ::    |     : | ||   || :    : ||:     :
Db     320 QN---GKVIENSQSNVDGVSGNSITQSFCKSQKTAFGDIDDFNKKGGLKQMGKALAQAMV 376

Qy     179 ----------------DSFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGC 221
                          ::| |  ||           || :      |||: | :
Db     377 LVMSIWDDHAANMLWLDSTYPVPKVPGA-------YRGSGPTTS----GVPAEVDANAPN 425

Qy     222 KRNDDSSFPV----FTPPSGGNGGT--------GTPT-STAPGSGQTSPGGGSGCTSQKW 268
             :   |:        :| |||: ||          :|| |||  |  ::    ||  :  |
Db     426 SKVAFSNIKFGHLGISPFSGGSSGTPPSNPSSSASPTSSTAKPSSTSTASNPSGTGAAHW 485

Qy     269 AQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           |||||||||| |||    || | :| ||||:
Db     486 AQCGGIGFSGPTTCPEPYTCAKDHDIYSQCV 516
```

RESULT 14
CEL1_AGABI
ID CEL1_AGABI STANDARD; PRT; 320 AA.
AC Q00023;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulose-growth-specific protein precursor.
GN CEL1.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D649;
RX MEDLINE=93012985; PubMed=1398098;
RA Raguz S., Yaguee E., Wood D.A., Thurston C.F.;
RT "Isolation and characterization of a cellulose-growth-specific gene
RT from Agaricus bisporus.";
RL Gene 119:183-190(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D649;
RX MEDLINE=94237428; PubMed=8181702;
RA Armesilla A.L., Thurston C.F., Yaguee E.;
RT "CEL1: a novel cellulose binding protein secreted by Agaricus
RT bisporus during growth on crystalline cellulose.";
RL FEMS Microbiol. Lett. 116:293-299(1994).
CC -!- FUNCTION: PROBABLE GLYCOSYL HYDROLASE ACTIVE ON CELLULOSE.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -!- SIMILARITY: BELONGS TO FAMILY 61 OF GLYCOSYL HYDROLASES.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M86356; AAA53434.1; -.
DR   HSSP; P00725; 2CBH.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR005103; Glyco_hydro_61.
DR   Pfam; PF00734; CBM_1; 1.
DR   Pfam; PF03443; Glyco_hydro_61; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30    320       CELLULOSE-GROWTH-SPECIFIC PROTEIN.
FT   DOMAIN       30    261       CATALYTIC (POTENTIAL).
FT   DOMAIN      262    285       LINKER (POTENTIAL).
FT   DOMAIN      286    320       CELLULOSE-BINDING (POTENTIAL).
FT   DISULFID    292    309       BY SIMILARITY.
FT   DISULFID    303    319       BY SIMILARITY.
FT   CARBOHYD    163    163       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   320 AA;  33754 MW;  60E2C8080895CA2B CRC64;

  Query Match             10.4%;  Score 172;  DB 1;  Length 320;
  Best Local Similarity   24.8%;  Pred. No. 6.6e-06;
  Matches   85;  Conservative   36;  Mismatches  132;  Indels   90;  Gaps   15;

Qy      10 TLAAALPLVASAASGSGQSTRY-WDCC-----KPSCAWPGKAAVSQPVYACDANFQRLSD 63
           | : || | |:     |    | ||        |     |: :: :|     | |  : |
Db      15 TFSLALGLFAAKVQAHGGVIGYSWDGTWYEGWHPYNTPVGQTSIERPW----ATFDPIMD 70

Qy      64 FNVQS-GCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVA 122
              : |||        |  |   |   |   | |:|       |      :|   |
Db      71 ATASTVGCNN-------DGNP-GPNQLTATVAAGTAITAYWNQVWPHPYGPMTTYLGKCP 122

Qy     123 GKTMVVQSTST------------GGDLG------------SNQFDIAMPGGGVGIFNGCS 158
           | :    :|::             | :|            :| :   :|
Db     123 GSSCDGVNTNSLKWFKIDEAGLLSGTVGKGVWGSGKMIDQNNSWTTTIP---------- 171

Qy     159 SQFGGLPGAQYGGISSRDQCDSFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCP--AEIV 216
               :|   |          | || : | |             | |   : |  :|:|
Db     172 ---STVPSGAYMIRFETIALHSLPAQIYPECA----------QLTITGGGNRAPTSSELV 218

Qy     217 ARSGCKRND--------------DSSFPVFTPP---SGGNGG--TGTPTSTAPGSGQTS 256
           :  |   |               |::: |  ||   ||||||  | || :| | :
Db     219 SFPGGYSNSDPGLTVNLYTQEAMTDTTYIVPGPPLYGSGGNGGSPTTTPHTTTPITTSPP 278

Qy     257 PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
           |    | |  :: ||||||::| | ||:   |: :|||||||
Db     279 PTSTPG-TIPQYGQCGGIGWTGGTGCVAPYQCKVINDYYSQCL 320


RESULT 15
GUN4_TRIRE
ID   GUN4_TRIRE     STANDARD;      PRT;   344 AA.
AC   O14405;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Endoglucanase IV precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase IV)
DE   (Cellulase IV) (EGIV).
GN   EGL4.
OS   Trichoderma reesei (Hypocrea jecorina).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Hypocreaceae; Hypocrea.
OX   NCBI_TaxID=51453;
RN   [1]
RP   SEQUENCE FROM N.A., ACTIVITY, AND INDUCTION.
RC   STRAIN=QM9414 / Rut C-30;
RX   MEDLINE=98036137; PubMed=9370370;
RA   Saloheimo M., Nakari-Setaelae T., Tenkanen M., Penttilae M.;
RT   "cDNA cloning of a Trichoderma reesei cellulase and demonstration of
RT   endoglucanase activity by expression in yeast.";
RL   Eur. J. Biochem. 249:584-591(1997).
CC   -!- FUNCTION: The biological conversion of cellulose to glucose
CC       generally requires three types of hydrolytic enzymes:
CC       (1) Endoglucanases which cut internal beta-1,4-glucosidic bonds;
CC       (2) Exocellobiohydrolases that cut the dissaccharide cellobiose
CC           from the nonreducing end of the cellulose polymer chain;
CC       (3) Beta-1,4-glucosidases which hydrolyze the cellobiose and other
CC           short cello-oligosaccharides to glucose.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC       linkages in cellulose.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- INDUCTION: By cellulose, cellobiose, lactose and sophorose.
CC   -!- PTM: May also be O-glycosylated.
CC   -!- SIMILARITY: BELONGS TO FAMILY 61 OF GLYCOSYL HYDROLASES.
CC   -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN
CC       (CBD).
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Y11113; CAA71999.1; -.
DR   HSSP; P00725; 2CBH.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR005103; Glyco_hydro_61.
DR   Pfam; PF00734; CBM_1; 1.
DR   Pfam; PF03443; Glyco_hydro_61; 1.
DR   ProDom; PD001821; CBD_fungal; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    344       ENDOGLUCANASE IV.
FT   DOMAIN       22    256       CATALYTIC (POTENTIAL).
FT   DOMAIN      257    307       LINKER (POTENTIAL).
FT   DOMAIN      308    344       CELLULOSE-BINDING (POTENTIAL).
FT   DISULFID    315    332       BY SIMILARITY.
FT   DISULFID    326    342       BY SIMILARITY.
FT   CARBOHYD     80     80       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    158    158       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   344 AA;  35510 MW;  7FBF1C4AB705350C CRC64;

  Query Match              9.6%;  Score 159;  DB 1;  Length 344;
  Best Local Similarity   24.1%;  Pred. No. 6.3e-05;
  Matches   71;  Conservative   31;  Mismatches  102;  Indels   90;  Gaps   14;

Qy      41 WPGKAAVSQPVYACDAN----------FQRLSDFNVQSGCNGGSAYS---CADQTPWAVN 87
           ||    :  :  |: :            | ::    : || : |:  |    ::   | |
Db     105 WPHPGPIVDYLANCNGDCETVDKTTLEFFKIDGVGLLSGGDPGTWASDVLISNNNTWVVK 164

Qy      88 --DNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMVVQSTSTGGDLGSNQ---- 141
             |||| |                  | |         :  ::   | |   |:
Db     165 IPDNLAPG----------------NYVL---------RHEIIALHSAGQANGAQNYPQC 198

Qy     142 FDIAMPGGGVGIFNGCSSQFGGLPGAQ-YGGISSRDQCDSFPAPLK---PGCQWRFDWFQ 197
           |:||: | |        | |  |: |   |          : : :||    ||
Db     199 FNIAVSGSG-------SLQPSGVLGTDLYHATDPGVLINIYTSPLNYIIPG--------- 242

Qy     198 NADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNGGTGTPTSTA-------- 249
               ||        |  :    |        |:    | | ||:| |   |:||
Db     243 ----PTVV---SGLPTSVAQGSSAATATASA----TVPGGGSGPTSRTTTTARTTQASSR 291

Qy     250 ----PGSGQTSPGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLNDYYSQCL 299
               | :  ::| ||   |   : |||| |:|| | |    ||  || ||:|||
Db     292 PSSTPPATTSAPAGGP--TQTLYGQCGGSGYSGPTRCAPPATCSTLNPYYAQCL 343
```

Search completed: February 24, 2003, 11:51:53
Job time : 16 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 24, 2003, 11:49:14 ; Search time 33 Seconds
(without alignments)
1866.912 Million cell updates/sec

Title: US-10-007-521-12
Perfect score: 1659
Sequence: 1 MRSTPVLRTTLAAALPLVAS.....TTCVSGTTCQKLNDYYSQCL 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1170	70.5	305	3	O93782	O93782 humicola gr
2	724	43.6	219	14	Q9JH92	Q9jh92 unclassifie
3	724	43.6	220	14	Q9JH84	Q9jh84 unclassifie
4	709	42.7	219	14	Q9JH91	Q9jh91 unclassifie
5	706.5	42.6	220	14	Q9JH83	Q9jh83 unclassifie
6	705	42.5	217	14	Q9JH95	Q9jh95 unclassifie
7	703	42.4	219	14	Q9JH86	Q9jh86 unclassifie
8	700	42.2	219	14	Q9JH87	Q9jh87 unclassifie
9	698.5	42.1	227	3	O93783	O93783 humicola gr
10	698	42.1	219	14	Q9JH90	Q9jh90 unclassifie
11	697.5	42.0	221	14	Q9JH94	Q9jh94 unclassifie
12	697	42.0	219	14	Q9JH89	Q9jh89 unclassifie
13	697	42.0	219	14	Q9JH88	Q9jh88 unclassifie
14	693	41.8	220	14	Q9JH96	Q9jh96 unclassifie
15	689.5	41.6	220	14	Q9JH93	Q9jh93 unclassifie
16	680.5	41.0	218	14	Q9JH85	Q9jh85 unclassifie
17	631	38.0	410	3	Q9P868	Q9p868 piromyces e
18	585.5	35.3	271	3	Q9UVP3	Q9uvp3 alternaria
19	543.5	32.8	242	5	O97401	O97401 phaedon coc
20	360.5	21.7	112	14	Q9JH99	Q9jh99 unclassifie
21	348.5	21.0	112	14	Q9JHA0	Q9jha0 unclassifie
22	347.5	20.9	112	14	Q9JH98	Q9jh98 unclassifie
23	338.5	20.4	112	14	Q9JHA1	Q9jha1 unclassifie
24	338.5	20.4	112	14	Q9JHA2	Q9jha2 unclassifie
25	330	19.9	111	14	Q9JH97	Q9jh97 unclassifie
26	247	14.9	353	3	Q9HE18	Q9he18 penicillium
27	215	13.0	514	3	O93832	O93832 trichoderma
28	200	12.1	505	3	Q9P8P3	Q9p8p3 trichoderma
29	191	11.5	525	3	Q12621	Q12621 humicola gr
30	188.5	11.4	302	3	Q99034	Q99034 trichoderma
31	183	11.0	523	3	Q9Y723	Q9y723 irpex lacte
32	181.5	10.9	529	3	Q8WZJ4	Q8wzj4 penicillium
33	176	10.6	283	3	Q96UV7	Q96uv7 lentinula e
34	169.5	10.2	536	3	Q9UVS8	Q9uvs8 aspergillus
35	163.5	9.9	536	3	Q9Y895	Q9y895 volvariella
36	163	9.8	412	3	O74169	O74169 aspergillus
37	161	9.7	397	3	Q8TFB0	Q8tfb0 trichoderma
38	161	9.7	504	3	Q01763	Q01763 phanerochae
39	161	9.7	516	3	Q96VU3	Q96vu3 lentinula e
40	160	9.6	517	3	Q9Y722	Q9y722 irpex lacte
41	156	9.4	856	3	O74170	O74170 aspergillus
42	152.5	9.2	290	3	Q9HEZ0	Q9hez0 phanerochae
43	150.5	9.1	526	3	Q9Y724	Q9y724 irpex lacte
44	150.5	9.1	860	3	Q02290	Q02290 neocallimas
45	150	9.0	479	3	Q9UUS4	Q9uus4 colletotric

ALIGNMENTS

RESULT 1
O93782
ID O93782 PRELIMINARY; PRT; 305 AA.
AC O93782;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Endoglucanase.
GN EGL3.
OS Humicola grisea var. thermoidea.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX NCBI_TaxID=5528;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO9854;
RX MEDLINE=99144540; PubMed=9990729;
RA Takashima S., Iikura H., Nakamura A., Hidaka M., Masaki H., Uozumi T.;
RT "Comparison of gene structures and enzymatic properties between two endoglucanases from Humicola grisea.";
RL J. Biotechnol. 67:85-97(1999).
DR EMBL; AB003107; BAA74956.1; -.
DR HSSP; P43316; 2ENG.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR000334; GH_45.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
SQ SEQUENCE 305 AA; 32174 MW; 28C979D6EDCD771D CRC64;

Query Match 70.5%; Score 1170; DB 3; Length 305;
Best Local Similarity 67.8%; Pred. No. 4e-85;
Matches 208; Conservative 44; Mismatches 45; Indels 10; Gaps 4;

```
Qy      1 MRSTPVLRTTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQR 60
          |||:|:| : : ||||::|| ||   |:|||||||||||| |  || |:|||::|:|||||
Db      1 MRSSPLLPSDVVAALPVLALAA--DGKSTRYWDCCKPSCGWAKKAPVNQPVFSCNANFQR 58
```

```
Qy      61 LSDFNVQSGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSG 119
           |:||: :|||  || ||||||||||||||: |:|||||||||| :|: |||||| ||||||
Db      59 LTDFDAKSGCEPGGVAYSCADQTPWAVNDDFAFGFAATSIAGSNEAGWCCACYELTFTSG 118

Qy     120 PVAGKTMVVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLPGAQYGGISSRDQCD 179
           ||||| |||||||||||||| ||: :|||||||||:||: ||||||| :|||||||::||
Db     119 PVAGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIFDGCTPQFGGLPGQRYGGISSRNECD 178

Qy     180 SFPAPLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFP-VFTPPSGG 238
            ||  ||||| |||||:|||||:|:|:|||||||:|||:||:|||| :|| |  | |
Db     179 RFPDALKPGCYWRFDWFKNADNPSFSFRQVQCPAELVARTGCRRNDDGNFPAVQIPSSST 238

Qy     239 NGGTGTPTSTAPGSGQTS------PGGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCQKLN 292
           :  | ||||:  |  |:      |   ||||:::|||||| |:|||||||:|:|| |:|
Db     239 SSPVGQPTSTSTTSTSTTSSPPVQPTTPSGCTAERWAQCGGNGWSGCTTCVAGSTCTKIN 298

Qy     293 DYYSQCL 299
           |:| |||
Db     299 DWYHQCL 305


RESULT 2
Q9JH92
ID   Q9JH92      PRELIMINARY;      PRT;   219 AA.
AC   Q9JH92;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Family 45 cellulase homologue.
OS   unclassified eukaryotes.
OC   Eukaryota.
OX   NCBI_TaxID=42452;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT   "Diverse genes of family 45 cellulase homologues of the symbiotic
RT   protists in the hindgut of termite Reticulitermes speratus.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB045171; BAA98041.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ   SEQUENCE   219 AA;  23001 MW;  5F2EB81A6DE926CE CRC64;

  Query Match             43.6%;  Score 724;  DB 14;  Length 219;
  Best Local Similarity   60.3%;  Pred. No. 6.2e-50;
  Matches  132;  Conservative   26;  Mismatches   53;  Indels    8;  Gaps    4;

Qy      11 LAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYAC--DANFQRLSDFNVQS 68
           |   | |: |    ||::|||||||| ||||  ||||:|||  |  |   :  |:  |:|
Db       2 LVFILALILSVFGDSGRTTRYWDCCKASCAWEKKAAVTQPVDTCGKDGTTRVASNDTVKS 61

Qy      69 GCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMVV 128
            |:||  | | || ||||||::|||||| :  || ||  || || |||||||| || |||
Db      62 ACDGGEGYMCYDQAPWAVNDSVAYGFAAAACCGG-ESGACCNCYELTFTSGPVNGKKMVV 120

Qy     129 QSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLP----GAQYGGISSRDQCDSFPAP 184
           | |:||||||||||:|:|||||||:|||: |  | |    |::|||:||| :|   |:
Db     121 QVTNTGGDLGSNQFDLAIPGGGVGIYNGCTQQ-SGAPADGWGSRYGGVSSRSECSQLPSG 179

Qy     185 LKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKR 223
           |: ||||||||||||||:  | || || |: |:: |||
Db     180 LQAGCQWRFDWFQNADNPSINFNQVTCPGELTAKTNCKR 218


RESULT 3
Q9JH84
ID   Q9JH84      PRELIMINARY;      PRT;   220 AA.
AC   Q9JH84;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Family 45 cellulase homologue.
OS   unclassified eukaryotes.
OC   Eukaryota.
OX   NCBI_TaxID=42452;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT   "Diverse genes of family 45 cellulase homologues of the symbiotic
RT   protists in the hindgut of termite Reticulitermes speratus.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB045179; BAA98049.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ   SEQUENCE   220 AA;  23274 MW;  26AF5357512EA061 CRC64;

  Query Match             43.6%;  Score 724;  DB 14;  Length 220;
  Best Local Similarity   59.8%;  Pred. No. 6.2e-50;
  Matches  131;  Conservative   30;  Mismatches   50;  Indels    8;  Gaps    4;

Qy      11 LAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYAC--DANFQRLSDFNVQS 68
           |  :| |::     ||::|||||||| |||| ||||:|||  |  |   :  |:  |:|
Db       3 LLFSLCLISWLVGDSGRTTRYWDCCKASCAWEKKAAVTQPVDTCGKDGTTRLASNDTVKS 62

Qy      69 GCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMVV 128
            |:||  | | || ||||||::|||||| :  || ||  || || |||||||| || |||
Db      63 SCDGGDGYMCYDQAPWAVNDSVAYGFAAAACCGG-ETGACCNCYELTFTSGPVNGKKMVV 121

Qy     129 QSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLP----GAQYGGISSRDQCDSFPAP 184
           | |:||||||||||:|:|||||||:|||: |  | |    |::|||:||| :|   |:
Db     122 QVTNTGGDLGSNQFDLAIPGGGVGIYNGCTQQ-SGAPADGWGSRYGGVSSRSECSQLPSG 180

Qy     185 LKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKR 223
           |: ||||||||||||||:  | || || |: |:: |||
Db     181 LQAGCQWRFDWFQNADNPSINFNQVTCPSELIARTNCKR 219


RESULT 4
Q9JH91
ID   Q9JH91      PRELIMINARY;      PRT;   219 AA.
AC   Q9JH91;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Family 45 cellulase homologue.
OS   unclassified eukaryotes.
OC   Eukaryota.
OX   NCBI_TaxID=42452;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT   "Diverse genes of family 45 cellulase homologues of the symbiotic
RT   protists in the hindgut of termite Reticulitermes speratus.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB045172; BAA98042.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ   SEQUENCE   219 AA;  23126 MW;  A712EF3F3CAB041C CRC64;

  Query Match             42.7%;  Score 709;  DB 14;  Length 219;
  Best Local Similarity   59.4%;  Pred. No. 9.5e-49;
  Matches  130;  Conservative   28;  Mismatches   53;  Indels    8;  Gaps    4;

Qy      11 LAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYAC--DANFQRLSDFNVQS 68
           |   | |::     ||::|||||||| |||| ||||:|||  |  |   :  |:  |:|
Db       2 LLFTLCLISWILGDSGRTTRYWDCCKASCAWEKKAAVTQPVDTCAKDGTTRVASNDTVKS 61

Qy      69 GCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMVV 128
            |:||  | | || ||||||::|||||| :  || ||  || || |||||||| || |||
```

```
Db     62 CCDGGEGYMCYDQAPWAVNDSVAYGFAAAACCGG-ESGACCNCYELTFTSGPVNGKKMVV 120

Qy    129 QSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLP----GAQYGGISSRDQCDSFPAP 184
          | |:||||||||||||:|:|||||||:|||: |  | |    |::|||:||| :|   |:
Db    121 QVTNTGGDLGSNQFDLAIPGGGVGIYNGCTQQ-SGAPSDGWGSRYGGVSSRSECSQLPSG 179

Qy    185 LKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKR 223
          |: |||| |||||||||:  | || ||:|: |:: |||
Db    180 LQAGCQWTFDWFQNADNPSINFNQVTCPSELTAKTNCKR 218


RESULT 5
Q9JH83
ID   Q9JH83      PRELIMINARY;      PRT;   220 AA.
AC   Q9JH83;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Family 45 cellulase homologue.
OS   unclassified eukaryotes.
OC   Eukaryota.
OX   NCBI_TaxID=42452;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT   "Diverse genes of family 45 cellulase homologues of the symbiotic
RT   protists in the hindgut of termite Reticulitermes speratus.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB045167; BAA98037.1; -.
DR   EMBL; AB045166; BAA98036.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ   SEQUENCE   220 AA;  23108 MW;  97738D831BCFA5F3 CRC64;

  Query Match             42.6%;  Score 706.5;  DB 14;  Length 220;
  Best Local Similarity   57.9%;  Pred. No. 1.5e-48;
  Matches  128;  Conservative   32;  Mismatches   52;  Indels    9;  Gaps    5;

Qy     11 LAAALPLVASAASG-SGQSTRYWDCCKPSCAWPGKAAVSQPVYAC--DANFQRLSDFNVQ 67
          |     |:||   | ||::|||||||| || |  || ||:|:  |  |   :  |:  |:
Db      2 LVFVFSLLASVLFGDSGKTTRYWDCCKGSCGWEAKADVSKPIDTCAKDGTTRVASNDTVK 61

Qy     68 SGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMV 127
          |||:||  | | ||||| |||: | |||| :|:|| | : || || |||||||| || |
Db     62 SGCDGGDGYMCYDQTPWGVNDSYALGFAAAAISGG-EKAACCNCYELTFTSGPVNGKKMT 120

Qy    128 VQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLP----GAQYGGISSRDQCDSFPA 183
          || |:|||||||||||:|:|||||||:|||::|  | |    |::|||:||| :|   |:
Db    121 VQVTNTGGDLGSNQFDLAIPGGGVGIYNGCTAQ-SGAPADGWGSRYGGVSSRSECSQLPS 179

Qy    184 PLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRN 224
          |: |||||||||||||:  |  | ||:|::|:: |:||
Db    180 GLQAGCQWRFDWFQNADNPSMNFNVVSCPSELIAKTNCRRN 220


RESULT 6
Q9JH95
ID   Q9JH95      PRELIMINARY;      PRT;   217 AA.
AC   Q9JH95;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Family 45 cellulase homologue.
OS   unclassified eukaryotes.
OC   Eukaryota.
OX   NCBI_TaxID=42452;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT   "Diverse genes of family 45 cellulase homologues of the symbiotic
RT   protists in the hindgut of termite Reticulitermes speratus.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB045168; BAA98038.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ   SEQUENCE   217 AA;  22796 MW;  660DD1346B3769DB CRC64;

  Query Match             42.5%;  Score 705;  DB 14;  Length 217;
  Best Local Similarity   58.7%;  Pred. No. 2e-48;
  Matches  125;  Conservative   33;  Mismatches   47;  Indels    8;  Gaps    4;

Qy     18 VASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYAC--DANFQRLSDFNVQSGCNGGSA 75
          :|||   ||::|||||||| || |  || ||:|:  |  |   :  |:  |:|||:||
Db      7 LASAFCDSGKTTRYWDCCKGSCGWEAKADVSKPIDTCAKDGTTRVASNDTVKSGCDGGDG 66

Qy     76 YSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMVVQSTSTGG 135
          : | ||||| |||: | |||| :|:|| | : || || |||||||| || | || |:|||
Db     67 FMCYDQTPWGVNDSYALGFAAAAISGG-EKAACCNCYELTFTSGPVNGKKMTVQVTNTGG 125

Qy    136 DLGSNQFDIAMPGGGVGIFNGCSSQFGGLP----GAQYGGISSRDQCDSFPAPLKPGCQW 191
          ||||||||:|:|||||||:|||::|  | |    |::|||:||| :|   |: |: ||||
Db    126 DLGSNQFDLAIPGGGVGIYNGCTAQ-SGAPADGWGSRYGGVSSRSECSQLPSGLQAGCQW 184

Qy    192 RFDWFQNADNPTFTFQQVQCPAEIVARSGCKRN 224
          ||||||||||:  |  | ||:|::|:: |:||
Db    185 RFDWFQNADNPSMNFNVVSCPSELIAKTNCRRN 217


RESULT 7
Q9JH86
ID   Q9JH86      PRELIMINARY;      PRT;   219 AA.
AC   Q9JH86;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Family 45 cellulase homologue.
OS   unclassified eukaryotes.
OC   Eukaryota.
OX   NCBI_TaxID=42452;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT   "Diverse genes of family 45 cellulase homologues of the symbiotic
RT   protists in the hindgut of termite Reticulitermes speratus.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB045177; BAA98047.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ   SEQUENCE   219 AA;  23158 MW;  ECD686EAB8ED1DD1 CRC64;

  Query Match             42.4%;  Score 703;  DB 14;  Length 219;
  Best Local Similarity   58.1%;  Pred. No. 2.9e-48;
  Matches  125;  Conservative   31;  Mismatches   51;  Indels    8;  Gaps    4;

Qy     15 LPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYAC--DANFQRLSDFNVQSGCNG 72
          |  |  : : ||::|||||||| || |  || | :|:  |  |   :  |:  |:|||:|
Db      6 LTFVGLSLAESGKTTRYWDCCKGSCGWEKKANVDKPIDTCAKDGTTRVASNDTVKSGCDG 65

Qy     73 GSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMVVQSTS 132
          |  | | |||||:|||: :|||||  :  || ||  || || |||||||| || |:|| |:
Db     66 GEGYMCYDQTPWSVNDSYSYGFAAAACCGG-ESGACCGCYDLTFTSGPVNGKHMIVQITN 124

Qy    133 TGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLP----GAQYGGISSRDQCDSFPAPLKPG 188
          ||||||||||:|:|||||||:|||::|  | |    |::|||:||| :|   |: |: |
Db    125 TGGDLGSNQFDLAIPGGGVGIYNGCTAQ-SGAPSDGWGSRYGGVSSRSECSQLPSGLQAG 183

Qy    189 CQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKR 223
          ||||||||||||||  |  |:|||||:|:: | |
Db    184 CQWRFDWFQNADNPNINFSSVRCPAEIIAKTNCNR 218
```

```
RESULT 8
Q9JH87
ID   Q9JH87      PRELIMINARY;      PRT;   219 AA.
AC   Q9JH87;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Family 45 cellulase homologue.
OS   unclassified eukaryotes.
OC   Eukaryota.
OX   NCBI_TaxID=42452;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT   "Diverse genes of family 45 cellulase homologues of the symbiotic
RT   protists in the hindgut of termite Reticulitermes speratus.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB045176; BAA98046.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ   SEQUENCE   219 AA;  23134 MW;  4BDEF4EC9ACC772D CRC64;

  Query Match             42.2%;  Score 700;  DB 14;  Length 219;
  Best Local Similarity   57.7%;  Pred. No. 4.9e-48;
  Matches  124;  Conservative   34;  Mismatches   49;  Indels    8;  Gaps    4;

Qy      15 LPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYAC--DANFQRLSDFNVQSGCNG 72
           |  :: : : ||::||||||| || |  || | :|:  |  |   :  |:  |:|||:|
Db       6 LAFISWSLADSGRTTRYWDCCKGSCGWEKKANVDKPIDTCAKDGTTRVASNDTVKSGCDG 65

Qy      73 GSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMVVQSTS 132
           |  : | ||||| |:|:|:||||| :  || ||  || || |||||||| || |||| |:
Db      66 GDGFMCYDQTPWQVSDSLSYGFAAAACCGG-ESGACCGCYELTFTSGPVNGKKMVVQITN 124

Qy     133 TGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLP----GAQYGGISSRDQCDSFPAPLKPG 188
           ||||||||||:|:|||||||:|||::|  | |    |::|||:||| :|   |: |: |
Db     125 TGGDLGSNQFDLAIPGGGVGIYNGCTAQ-SGAPSDGWGSRYGGVSSRSECSQLPSGLQAG 183

Qy     189 CQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKR 223
           |||||||||||||:  | || ||:||:|:: | |
Db     184 CQWRFDWFQNADNPSINFSQVSCPSEIIAKTHCNR 218


RESULT 9
O93783
ID   O93783      PRELIMINARY;      PRT;   227 AA.
AC   O93783;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Endoglucanase.
GN   EGL4.
OS   Humicola grisea var. thermoidea.
OC   Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX   NCBI_TaxID=5528;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IFO9854;
RX   MEDLINE=99144540; PubMed=9990729;
RA   Takashima S., Iikura H., Nakamura A., Hidaka M., Masaki H., Uozumi T.;
RT   "Comparison of gene structures and enzymatic properties between two
RT   endoglucanases from Humicola grisea.";
RL   J. Biotechnol. 67:85-97(1999).
DR   EMBL; AB003108; BAA74957.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ   SEQUENCE   227 AA;  24240 MW;  873553E76F5C39E4 CRC64;

  Query Match             42.1%;  Score 698.5;  DB 3;  Length 227;
  Best Local Similarity   52.7%;  Pred. No. 6.7e-48;
  Matches  137;  Conservative   26;  Mismatches   46;  Indels   51;  Gaps    8;

Qy       9 TTLAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYACDANFQRLSD-FNVQ 67
           |||   || :|:| ||||::|||||||||||||||||     ||  ||     | |  | :
Db       6 TTLLTLLPALAAAQSGSGRTTRYWDCCKPSCAWPGKGPA--PVRTCDRWDNPLFDGGNTR 63

Qy      68 SGCN-GGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTM 126
           |||: || || |:||:||||:|:||||:|| :||| :|  ||||| ||||||||||| |
Db      64 SGCDAGGGAYMCSDQSPWAVSDDLAYGWAAVNIAGSNERQWCCACYELTFTSGPVAGKRM 123

Qy     127 VVQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLP---GAQYGGISSRDQCDSFPA 183
           :||:::|||||:| |||||||||||||| |: |:|  |   | :|||| | :||:||
Db     124 IVQASNTGGDLGNNHFDIAMPGGGVGIFNACTDQYGAPPNGWGQRYGGISQRHECDAFPE 183

Qy     184 PLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRNDDSSFPVFTPPSGGNGGTG 243
            ||||| |||||               | :                 :| | |
Db     184 KLKPGCYWRFDW---------------CVS-----------------LFPPLS------- 204

Qy     244 TPTSTAPGSGQTSPGGGSGC 263
             |  ||:|||    |  |
Db     205 --LSLPPGTGQTM---GRSC 219


RESULT 10
Q9JH90
ID   Q9JH90      PRELIMINARY;      PRT;   219 AA.
AC   Q9JH90;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Family 45 cellulase homologue.
OS   unclassified eukaryotes.
OC   Eukaryota.
OX   NCBI_TaxID=42452;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT   "Diverse genes of family 45 cellulase homologues of the symbiotic
RT   protists in the hindgut of termite Reticulitermes speratus.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB045173; BAA98043.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ   SEQUENCE   219 AA;  23037 MW;  372E016415530A9A CRC64;

  Query Match             42.1%;  Score 698;  DB 14;  Length 219;
  Best Local Similarity   56.7%;  Pred. No. 7.1e-48;
  Matches  122;  Conservative   36;  Mismatches   49;  Indels    8;  Gaps    4;

Qy      15 LPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYAC--DANFQRLSDFNVQSGCNG 72
           |  :  : : ||::||||||| || |  || | :|:  |  |   :  |:  |:|||:|
Db       6 LTFIGLSLADSGKTTRYWDCCKGSCGWEKKANVDKPIDTCAKDGTTRVASNDTVKSGCDG 65

Qy      73 GSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMVVQSTS 132
           |: : | ||||| |:|:|:||||| :  || ||  || || |||||||| || |:|| |:
Db      66 GTGFMCYDQTPWQVSDSLSYGFAAAACCGG-ESGACCGCYELTFTSGPVNGKKMIVQITN 124

Qy     133 TGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLP----GAQYGGISSRDQCDSFPAPLKPG 188
           ||||||||||:|:|||||||:|||::|  | |    |::|||:||| :|   |: |: |
Db     125 TGGDLGSNQFDLAIPGGGVGIYNGCTAQ-SGAPSDGWGSRYGGVSSRSECSQLPSGLQAG 183

Qy     189 CQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKR 223
           |||||||||||||:  |  |:|||:||:|:: | |
Db     184 CQWRFDWFQNADNPSINFSNVKCPSEIIAKTNCNR 218


RESULT 11
Q9JH94
ID   Q9JH94      PRELIMINARY;      PRT;   221 AA.
```

```
AC   Q9JH94;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Family 45 cellulase homologue.
OS   unclassified eukaryotes.
OC   Eukaryota.
OX   NCBI_TaxID=42452;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT   "Diverse genes of family 45 cellulase homologues of the symbiotic
RT   protists in the hindgut of termite Reticulitermes speratus.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB045169; BAA98039.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ   SEQUENCE   221 AA;  23220 MW;  8A84CEB0A8C46372 CRC64;

  Query Match           42.0%;  Score 697.5;  DB 14;  Length 221;
  Best Local Similarity  56.9%;  Pred. No. 7.9e-48;
  Matches  124;  Conservative   37;  Mismatches   48;  Indels    9;  Gaps    5;

Qy      14 ALPLVASAASG-SGQSTRYWDCCKPSCAWPGKAAVSQPVYAC--DANFQRLSDFNVQSGC 70
           |  |: :|  | ||::|||||||| || |  || ||:|:  |  |   :  |:  |:||| 
Db       5 AFVLLLNAIFGDSGKTTRYWDCCKGSCGWEAKADVSKPIDTCAKDGTTRVASNDTVKSGC 64

Qy      71 NGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMVVQS 130
           :||  : | |||||||||: : |||| :::|| | : || || |||||||| || : ||
Db      65 DGGDGFMCYDQTPWAVNDSYSLGFAAAAVSGG-EKAACCQCYELTFTSGPVNGKKLTVQV 123

Qy     131 TSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLP----GAQYGGISSRDQCDSFPAPLK 186
           |:|||||||||||:|:|||||||:|||::|  | |    |::|||:||| :|   |: |: 
Db     124 TNTGGDLGSNQFDLAIPGGGVGIYNGCTAQ-SGAPADGWGSRYGGVSSRSECSQLPSGLQ 182

Qy     187 PGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKRN 224
           ||||||||||||||:  |  | ||:|:::::: ||||
Db     183 AGCQWRFDWFQNADNPSINFNVVSCPSELISKTNCKRN 220


RESULT 12
Q9JH89
ID   Q9JH89      PRELIMINARY;      PRT;   219 AA.
AC   Q9JH89;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Family 45 cellulase homologue.
OS   unclassified eukaryotes.
OC   Eukaryota.
OX   NCBI_TaxID=42452;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT   "Diverse genes of family 45 cellulase homologues of the symbiotic
RT   protists in the hindgut of termite Reticulitermes speratus.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB045174; BAA98044.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ   SEQUENCE   219 AA;  23033 MW;  CA295CAD8F393199 CRC64;

  Query Match           42.0%;  Score 697;  DB 14;  Length 219;
  Best Local Similarity  57.2%;  Pred. No. 8.5e-48;
  Matches  123;  Conservative   34;  Mismatches   50;  Indels    8;  Gaps    4;

Qy      15 LPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYAC--DANFQRLSDFNVQSGCNG 72
           |  :  : : ||::|||||||| || |  || | :|:  |  |   :  |:  |:|||:| 
Db       6 LTFIGLSLADSGKTTRYWDCCKGSCGWEKKANVDKPIDTCAKDGVTRVASNDTVKSGCDG 65

Qy      73 GSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMVVQSTS 132
           |  : | ||||| |:|:|:||||| :  || ||  || || |||||||| || |:|| |: 
Db      66 GDGFMCYDQTPWQVSDSLSYGFAAAACCGG-ESGACCGCYELTFTSGPVNGKKMIVQITN 124

Qy     133 TGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLP----GAQYGGISSRDQCDSFPAPLKPG 188
           |||||||||||:|:|||||||:|||::|  | |    |::|||:||| :|   |: |: | 
Db     125 TGGDLGSNQFDLAIPGGGVGIYNGCTAQ-SGAPSDGWGSRYGGVSSRSECSQLPSGLQAG 183

Qy     189 CQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKR 223
           |||||||||||||:  |  |:|||||:|:: | |
Db     184 CQWRFDWFQNADNPSINFSNVKCPAEIIAKTNCNR 218


RESULT 13
Q9JH88
ID   Q9JH88      PRELIMINARY;      PRT;   219 AA.
AC   Q9JH88;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Family 45 cellulase homologue.
OS   unclassified eukaryotes.
OC   Eukaryota.
OX   NCBI_TaxID=42452;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT   "Diverse genes of family 45 cellulase homologues of the symbiotic
RT   protists in the hindgut of termite Reticulitermes speratus.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB045175; BAA98045.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ   SEQUENCE   219 AA;  23030 MW;  179BF1344C6D7024 CRC64;

  Query Match           42.0%;  Score 697;  DB 14;  Length 219;
  Best Local Similarity  56.6%;  Pred. No. 8.5e-48;
  Matches  124;  Conservative   33;  Mismatches   54;  Indels    8;  Gaps    4;

Qy      11 LAAALPLVASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYAC--DANFQRLSDFNVQS 68
           |   |  :  : : ||::|||||||| || |  || | :|:  |  |   :  |:  |:| 
Db       2 LLQILTFIGLSLAESGKTTRYWDCCKGSCGWEKKANVDKPIDTCAKDGTTRVASNDTVKS 61

Qy      69 GCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMVV 128
           ||:||: | | ||||| |:|:|:||||| :  || ||  || || |||||||| || |:| 
Db      62 GCDGGTGYMCYDQTPWQVSDSLSYGFAAAACCGG-ESGACCGCYELTFTSGPVNGKKMIV 120

Qy     129 QSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLP----GAQYGGISSRDQCDSFPAP 184
           | |:|||||||||:|:|||||||:|||::|  | |    |::|||:||| :|   |: 
Db     121 QITNTGGDLGSNQFDLAIPGGGVGIYNGCTSQ-SGAPADGWGSRYGGVSSRSECSQLPSG 179

Qy     185 LKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKR 223
           |: |||||||| |||||   |  |:||:||:|:: | |
Db     180 LQAGCQWRFDWFANADNPNINFTNVKCPSEIIAKTNCNR 218


RESULT 14
Q9JH96
ID   Q9JH96      PRELIMINARY;      PRT;   220 AA.
AC   Q9JH96;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Family 45 cellulase homologue.
OS   unclassified eukaryotes.
OC   Eukaryota.
OX   NCBI_TaxID=42452;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
```

```
RT   "Diverse genes of family 45 cellulase homologues of the symbiotic
RT   protists in the hindgut of termite Reticulitermes speratus.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB045165; BAA98035.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ   SEQUENCE   220 AA;  23148 MW;  4F0652F8E8D269D5 CRC64;

  Query Match             41.8%;  Score 693;  DB 14;  Length 220;
  Best Local Similarity   58.7%;  Pred. No. 1.8e-47;
  Matches  125;  Conservative   30;  Mismatches   50;  Indels    8;  Gaps    4;

Qy      18 VASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYAC--DANFQRLSDFNVQSGCNGGSA 75
           :||    ||::|||||||| || |  || ||:|:  |  |   :  |:  |:|||:||
Db      10 LASVFGESGRTTRYWDCCKGSCGWEAKADVSKPIDTCAKDGTTRVASNDTVKSGCDGGEG 69

Qy      76 YSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMVVQSTSTGG 135
           | | |||| |||: | |||| :|:|| | : || || |||||||| || | || |:|||
Db      70 YMCYDQTPRAVNDSYAIGFAAAAISGG-EKAACCQCYELTFTSGPVNGKKMTVQVTNTGG 128

Qy     136 DLGSNQFDIAMPGGGVGIFNGCSSQFGGLP----GAQYGGISSRDQCDSFPAPLKPGCQW 191
           ||||||||:|:|||||||:|||::|  | |    |::|||:||| :|   |: |: ||||
Db     129 DLGSNQFDLAIPGGGVGIYNGCTAQ-SGAPADGWGSRYGGVSSRSECSQLPSGLQAGCQW 187

Qy     192 RFDWFQNADNPTFTFQQVQCPAEIVARSGCKRN 224
           ||||||||||:  |  | || |::| : |:||
Db     188 RFDWFQNADNPSINFNVVSCPGELIANTNCRRN 220


RESULT 15
Q9JH93
ID   Q9JH93      PRELIMINARY;      PRT;   220 AA.
AC   Q9JH93;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Family 45 cellulase homologue.
OS   unclassified eukaryotes.
OC   Eukaryota.
OX   NCBI_TaxID=42452;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT   "Diverse genes of family 45 cellulase homologues of the symbiotic
RT   protists in the hindgut of termite Reticulitermes speratus.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB045170; BAA98040.1; -.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF02015; Glyco_hydro_45; 1.
DR   PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ   SEQUENCE   220 AA;  23091 MW;  35F174F9B79C2D65 CRC64;

  Query Match             41.6%;  Score 689.5;  DB 14;  Length 220;
  Best Local Similarity   56.4%;  Pred. No. 3.4e-47;
  Matches  124;  Conservative   37;  Mismatches   50;  Indels    9;  Gaps    5;

Qy      11 LAAALPL-VASAASGSGQSTRYWDCCKPSCAWPGKAAVSQPVYAC--DANFQRLSDFNVQ 67
           ||  : | :||    :|::|||||||| || |  || | :|:  |  |   :  |:  |:
Db       2 LAFVITLFLASVFGDTGRTTRYWDCCKGSCGWEKKANVDKPIDTCAKDGTTRVASNDTVK 61

Qy      68 SGCNGGSAYSCADQTPWAVNDNLAYGFAATSIAGGSESSWCCACYALTFTSGPVAGKTMV 127
           |||:||  | | ||:|| |||: | |||| :::|| ||: || || |||||||| || |
Db      62 SGCDGGDGYMCYDQSPWGVNDSFALGFAAAAVSGG-ESAACCNCYELTFTSGPVNGKKMT 120

Qy     128 VQSTSTGGDLGSNQFDIAMPGGGVGIFNGCSSQFGGLP----GAQYGGISSRDQCDSFPA 183
           || |:||||||||||:|:|||||||:|||::|  | |    |::|||:||| :|   |:
Db     121 VQVTNTGGDLGSNQFDLAIPGGGVGIYNGCTAQ-SGAPADGWGSRYGGVSSRSECCQLPS 179

Qy     184 PLKPGCQWRFDWFQNADNPTFTFQQVQCPAEIVARSGCKR 223
           |: |||||||||||||: :|  | ||:|::|:: |:|
Db     180 GLQAGCQWRFDWFQNADNPSISFNVVSCPSELIAKTNCRR 219


Search completed: February 24, 2003, 11:52:33
Job time : 35 secs
```